Translating genetic findings on chromosome 10q24 into an understanding of neurobiological risk mechanisms for schizophrenia

Rafagnin Duarte, Rodrigo Roberto

Awarding institution:
King's College London

Translating genetic findings on chromosome 10q24 into an understanding of neurobiological risk mechanisms for schizophrenia

Rodrigo Roberto Rafagnin Duarte

Thesis submitted for the degree of Doctor of Philosophy

Department of Basic & Clinical Neuroscience
Institute of Psychiatry, Psychology & Neuroscience
King's College London

Submitted in September, 2016

Abstract

Schizophrenia is a poorly understood mental disorder, for which treatment options show variable efficacy and serious side effects. The aetiology of schizophrenia remains unclear, but genetics are known to make a large contribution to risk. It is expected that a third to half of the genetic variation contributing to risk is accounted for by common genetic variants. So far, large-scale genome-wide association studies (GWAS) have implicated 108 common genome-wide significant loci in susceptibility to schizophrenia. While analysis of larger cohorts will likely identify more common variants involved in susceptibility, the functional characterisation of individual risk loci is necessary to identify the risk genes and the functional effects associated with risk variation, potentially exposing novel drug targets. The functional characterisation of a risk locus, on chromosome 10q24, is presented in this thesis. As the best supported risk variants at the locus (rs11191419 and chr10_104957618_I) are non-coding, risk alleles were hypothesised to alter the regulation of one or more genes in the region. Measures of allele-specific expression were used to investigate *cis*-regulatory effects associated with the risk variants on the primary positional candidates *BORCS7, AS3MT, CNNM2,* and *NT5C2* in the human brain. The risk allele of rs11191419 was found to be associated with increased allelic expression of *BORCS7* and *AS3MT,* and with decreased expression of *NT5C2.* The risk allele of chr10_104957618_I was associated with decreased expression of *BORCS7*, *AS3MT,* and *NT5C2*. A RNA-sequencing pilot study was performed to identify transcripts produced by these genes in brain tissues where *cis*-regulatory effects associated with risk alleles were observed. This revealed RNA expression of RefSeq transcripts of *BORCS7*, *AS3MT,* and *NT5C2*, as well as of novel transcripts of *AS3MT* and *NT5C2*. Using immunohistochemistry, the cytosolic 5'-nucleotidase II, produced by *NT5C2,* was found to co-localise with neurons, glial cells and neuropil in the adult dorsolateral prefrontal cortex (DLPFC). Knockdown of *NT5C2* in neural progenitor cells was found to alter expression of genes involved in the regulation of the cytoskeleton, cellular metabolism and AMPK signalling. These results suggest neurobiological mechanisms through which genetic variation on chromosome 10q24 confers risk to schizophrenia.

Acknowledgements

I would like to thank the following people who contributed intellectually, professionally (and emotionally) to this project:

- My Supervisors Dr Deepak Srivastava, Dr Nicholas Bray and Prof Robin Murray for the discussions, advice and opportunities given to me. You are absolutely fantastic researchers and excellent role models.

- My overseas sponsoring programme, Science without Borders (CAPES, Brazil), which provided me a full scholarship and the opportunity to study at King's, a world-leading research institution. A special thanks to the responsible person for my grant, Ms Fernanda de Arruda, for the assistance throughout the past years.

- My mother Mari, my sister Gisele, and my partner Tim, who are the source of my happiness, strength and drive for a prosperous life and career. I would also like to thank my beloved friends Alex, Eddy, Doug, Kelly, Caio, Jake, Ju, Pri, Kako and Vivi.

- My good friends Dr Fabio Vigil and Dr Robert Westphal and their beloved ones, Mrs Livia Ferreira and Ms Thao Le, for their friendship, support and beers at The Sun.

- All current and former members of the Bray's lab, Dr Katherine Navarrete, Ms Stephanie Robinson (and the little Imogen, our honorary lab member!), Ms Gemma McLaughlin, Dr Greg Anderson, Ms Carolina Toste, Mr Darren Byrne, who were all amazing lab mates and friends. I am extremely happy to have had the chance to work closely with all of you during my PhD.

- All current and former Srivastava's lab members, Ms Pooja Raval, Dr Carole Shum, Dr Michael Deans, Dr Katherine Sellers, Mr Filippo Erli, Mr Nick Gatford, Mr Iain Watson, Mr Alish Palmos, who were absolutely incredible lab mates, always helping each other out, and always knowing how to celebrate our achievements!

- Our collaborators at the Brain Bank, Dr Claire Troakes, Mr Mathew Nolan, Ms Sashika Selvackadunco and Dr Teresa Rodriguez, who contributed a lot of their time and expertise for this project to happen successfully.

- Current and former academic staff and students from the Behaviour Cell Unit who somehow contributed to this project by providing their time, resources or expertise, especially Dr Marie-Caroline Côtel, Dr Hemanth Nelvagal, Mr Robert Chesters, Dr Leo Perfect, Dr Marta Tarczyluk, Dr Aleksandra Maruszak, Dr Sandrine Thuret, Dr Anthony Vernon, Dr Brenda Williams, Prof Jon Cooper, Prof Jack Price, Prof Peter Giese, Dr Keiko Mizuno, Dr Salvatore Adinolfi.

- The incredible supporting staff from the Department of Basic & Clinical Neuroscience, Mrs Kate Grant, Mr Mario Mazzantini, Mr Jeremy Wiltshire, Ms Samantha Smith, Mrs Toni Moreby, Dr Rebecca Gresham.

- All King's College London staff, especially from the Education Support Team and King's Worldwide Department, Mr Benjamin Harrison, Mr William Fitzmaurice, Ms Dafina Shabani and Mr Oliver Trumble.

- Our collaborators at the MRC Social, Genetic & Developmental Psychiatry Centre at King's College London, Dr Timothy Powell, Dr Gerome Breen and Mr Sang Lee, who provided their time and expertise for the global gene expression analyses.

The work presented in this thesis was funded by a Science without Borders scholarship to the author (CAPES, Brazilian Ministry of Education, BEX 1279-13-0), a Medical Research Council grant to NJ Bray (#G0802166), a Wellcome Trust ISSF Grant to DP Srivastava (#097819). The human foetal material was provided by the Joint MRC/Wellcome Trust Human Developmental Biology Resource (#099175/Z/12/Z). Adult tissue samples were supplied by The London Neurodegenerative Diseases Brain Bank, which receives funding from the MRC and, as part of the Brains for Dementia Research programme, jointly funded by Alzheimer's Research UK and Alzheimer's Society.

Statement of Work

- **Chapter 1:** I wrote all parts of this chapter.
- **Chapter 2:** This chapter was published on the American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 2016, 171(6):806-814. I carried out all experimental procedures, statistical analyses and writing of this chapter.
- **Chapter 3:** Unpublished processed expression data from neuronal differentiation in cortical and hippocampal neural progenitor cells were provided by Dr Timothy Powell, Dr Sandrine Thuret, Dr Gerome Breen, Dr Deepak Srivastava, Dr Nick Bray and Dr Greg Anderson. I carried out all other experimental procedures, statistical analyses and writing of this chapter.
- **Chapter 4:** The neural progenitors derived from human induced pluripotent stem cells (hiPSCs) were kindly provided by Dr Carole Shum (Srivastava's lab) in collaboration with Prof Price's group. Immunohistochemistry was advised by Dr Marie-Caroline Côtel (Vernon's lab) and Dr Claire Troakes (MRC Brain Bank). Immunohistochemistry images were obtained with the assistance of Dr Hemanth Nelvaga and Dr Marta Tarczyluk (Cooper's lab). I carried out all experimental procedures, statistical analyses and writing of this chapter.
- **Chapter 5:** The microarray run performed in this chapter was carried out by Mr Sanghyuck Lee, from the BRC IoPPN Genomics & Biomarker Core Facility at the Social, Genetic & Developmental Psychiatry Centre (SGDP), King's College London. The microarray and gene ontology analyses were assisted by Dr Timothy Powell and Dr Gerome Breen from the SGDP. I carried out all other experimental procedures, statistical analyses and writing of this chapter.
- **Chapter 6:** I wrote all parts of this chapter.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 5

List of Abbreviations

- 3C: Chromosome conformation capture
- *ACTB:* Beta-actin
- AMY: Amygdala
- *AS3MT*: Arsenite methyltransferase
- *ATG4B*: Autophagy related 4B cysteine peptidase
- *B2M*: Beta-2-microglobulin
- BCA: Bicinchoninic acid
- BDNF: Brain-derived neurotrophic factor
- *BORCS7*: BLOC-1 related complex subunit 7
- *C4*: Complement 4
- CBC: Cerebral cortex
- cDNA: Complementary DNA
- *cN-II*: Cyclin M2
- *CNNM2*: Cyclin M2
- CNVs: Copy number variants
- *COMT*: Catechol-O-methyltransferase gene
- CREs: *Cis*-regulatory elements
- CTCF: Corrected total cell fluorescence
- DAB: Diaminobenzidine
- *DAOA*: D-amino acid oxidase activato
- ddNTPs: 2',3'-dideoxynucleotides
- *DISC1*: Disrupted in schizophrenia 1
- DLPFC: Dorsolateral prefrontal cortex
- DNA: Deoxyribonucleic acid
- dNTPs: Deoxynucleotides
- dsRNAs: Double stranded RNA
- *DTNBP1*: Dystrobrevin-binding protein
- ER: Endoplasmic reticulum
- ExoI: Exonuclease I
- fMRI: Functional magnetic resonance imaging
- FISH: Fluorescence *in situ* hybridisation
- FOV: Fields of view
- FPKM: Fragments per kilobase of exon per million fragments mapped
- GABA: Gamma-Aminobutyric acid
- *GAD1*: Glutamic acid decarboxylase 67
- *GAPDH*: Glyceraldehyde-3-phosphate dehydrogenase
- gDNA: Genomic DNA
- GFP: Green fluorescent protein

- GO: Gene Ontology
- GTEX: Genotype-Tissue Expression
- GWAS: Genome-wide association studies
- HIP: Hippocampus
- hiPSC: Human induced pluripotent stem cells
- HLA: Human leukocyte antigen
- *HNRNPA1*: Heterogeneous nuclear ribonucleoprotein A1
- IF-gamma: Interferon-gamma
- IL: Interleukin
- IMS: Industrial methylated spirit
- iPSC: Induced pluripotent stem cells
- KEGG: Kyoto Encyclopedia of Genes and Genomes
- LD: Linkage disequilibrium
- LSD: Lysergic acid diethylamide
- MAF: Minimum allele frequency
- *MAPK*: Mitogen-activated protein kinases
- MD: Mediodorsal nucleus of thalamus
- MHC: Major histocompatibility complex
- Min: Minute
- mQTL: Methylation quantitative trait loci
- MRI: Magnetic resonance imaging
- mRNA: Messenger RNA
- MWAS: Methylome-wide association studies
- NCX: Neocortex
- NGS: Next-generation sequencing OR normal goat serum
- *NMDAR*: Nmethyl-d-aspartic acid-type receptors
- NPC: Neural progenitor cell
- *NRG1*: Neuregulin 1
- *NT5C2*: Cytosolic 5'-nucleotidase II
- PBS: Phosphate-buffered saline
- PCP: Phencyclidine
- PCR: Polymerase chain reaction
- PSD95: Postsynaptic density protein 95
- *PSMC4*: Proteasome 26S subunit, ATPase 4
- qPCR: Quantitative polymerase chain reaction
- RACE: Rapid amplification of cDNA ends
- *RELN*: Reelin
- RIN: RNA Integrity number
- RISC: RNA Induced Silencing Complex

- RNA: Ribonucleic acid
- RNAi: RNA interference
- RNA-seq: RNA-sequencing
- ROI: Regions of interest
- *RPL13A*: ribosomal protein L13a
- *RPL30*: Ribosomal protein L30
- rRNA: Ribosomal RNA
- rSAP: Rapid shrimp alkaline phosphatase
- RT-qPCR: Reverse transcription quantitative polymerase chain reaction
- *SDHA*: Succinate Dehydrogenase complex, subunit A
- SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis
- Sec: Second
- *SETD1A*: SET domain containing 1A
- siRNA: Small interfering RNA
- SNP: Single nucleotide polymorphism
- SNV: Single nucleotide variants
- STR: Striatum
- *TAAR1*: Trace amine-associated receptor 1
- *TBCA*: Tubulin-specific chaperone A
- TBS: Tris-buffered saline
- *TGF*: Transforming growth factor
- TWAS: Transcriptome-wide association studies
- *UBC*: Ubiquitin C
- UTR: Untranslated region
- v/v: Volume by volume
- VNTR: Variable number tandem repeat
- *ZNF804A*: Zinc finger protein 804A

Cats drawn by the English artist Louis Wain (1860-1939) during his descent to mental illness. Experts believe that he suffered from a psychiatric condition, possibly schizophrenia, which is curiously reflected in his art.

Chapter 1

General Introduction

1.1. Summary

Schizophrenia is a heterogeneous psychiatric disorder characterised by distorted emotional and cognitive processes. The class of drugs currently used for treatment, termed antipsychotics, was introduced over 60 years ago and is frequently associated with poor prognostics due to variable efficacy and severe side effects. Neurobiological research of schizophrenia has provided broad clues to its aetiology. A brief summary of such a vast literature is presented in this chapter, which include findings from studies of functional and structural neuroimaging, pathophysiology, neurochemistry and epigenetics. In its capacity to identify molecules and genes directly implicated in disease, genetic investigation revolutionised schizophrenia research. With the advent of 'hypothesis-free', large-scale genome-wide association studies (GWAS) and sequencing projects led by large international consortia, however, schizophrenia research has been taken to a whole new level. In the 'post-GWAS Era', nevertheless, many obstacles hinder the interpretation of association signals, including linkage disequilibrium (LD), and tissue, time-dependent or long-range *cis*-regulatory effects. Examples of how to circumvent these problems are discussed, which include coupling expression data with genotypic information using rather sophisticated genome-wide approaches, or more cost-effective, gene-specific methods. At the end of the chapter, a brief introduction is given to the genome-wide significant association signal on chromosome 10q24, the third most significant locus associated with schizophrenia in the most recent GWAS, for which the functional characterisation is the main objective of this thesis. The focal positional candidate genes at this locus are *BORCS7, AS3MT, CNNM2,* and *NT5C2*, all of which have putative roles implicated in neural function or neurodevelopment.

1.2. Schizophrenia

Schizophrenia is a heterogeneous psychiatric disorder characterised by distorted cognitive and emotional processes that cause a combination of 'positive', 'negative' and cognitive symptoms (Owen et al., 2016). Psychosis, the mostly acknowledged feature of schizophrenia, is a condition described by the term positive symptoms, which include the presence of delusions; the maintenance of beliefs despite contradictory reality or rational arguments; and hallucinations, which refers to sensing events that are not physically present. The negative symptoms of schizophrenia, on the other hand, refer to the lack of motivation or self-drive (termed 'avolition'), as well as the social withdrawal observed in patients. Cognitive domains such as working memory, word fluency and decision-making are also frequently impaired (Gur and Gur, 2010; Wong and Van Tol, 2003). These symptoms occur in heterogeneous combinations in patients (Joyce and Roiser, 2007), and therefore it is not rare to refer to this disorder as 'the schizophrenias' (Peralta and Cuesta, 2011).

The life-time prevalence associated with schizophrenia ranges from 0.30%-0.66% in the world population (McGrath et al., 2008), which makes it a relatively common disease. Symptoms generally appear during early adulthood or late adolescence, causing a financial burden to the public health system, patients, families and society in general. The disease is associated with unemployment rates as high as 90% (Kooyman et al., 2007; Mangalore and Knapp, 2007; Marwaha and Johnson, 2004), decreased life expectancy, and is estimated to cost £11.8 billion per year only in England (Schizophrenia Commission, 2012). While no molecular or neuroimaging biomarkers have been yet identified, diagnosis is determined by medical interviewing. Guidelines for the diagnosis as per the latest version of the Diagnostic and Statistical Manual of Mental Disorders DSM-V (American Psychiatric Association, 2013) specify that the schizophrenia patient must present (1) two or more of the following symptoms: delusions, hallucinations, disorganised speech, grossly disorganised or catatonic behaviour, negative symptoms; (2) social/occupational dysfunction in

terms of work, interpersonal relations or self-care; (3) continuous sign of disturbances for over six months; (4) symptoms that are not associated with schizoaffective or mood disorders, or drug abuse.

Current treatment options involve family intervention and cognitive behaviour therapy in conjunction with medication (Pilling et al., 2002). The drugs prescribed, termed antipsychotics, however, are known to mostly affect dopaminergic transmission and thus mainly improve positive symptoms (the psychosis component), showing variable efficacy and serious side effects, while negative, cognitive and affective impairments generally remain unaltered (Moller and Czobor, 2015). Antipsychotics are classically divided in typical (first generation) and atypical (second generation) drugs, although a recent study found no robust evidence to clinically support this distinction (Leucht et al., 2009). Side effects associated with typical antipsychotics such as fluphenazine, haloperidol and chlorpromazine, and atypical antipsychotics such as ziprasidone, risperidone and clozapine, generally include extrapyramidal effects, weight gain and sedation (Leucht et al., 2012), with some of these drugs tending to cause a few specific side effects more than others. Approximately a third of patients do not respond to the first prescribed treatment (Hasan et al., 2012), and another third does not respond to treatment at all (Ackenheil and Weber, 2004). In this context, understanding the molecular underpinnings of this disorder has the potential of exposing novel drug targets and, possibly, result in more effective disease management (Marino et al., 2008). As it will be discussed later, genetic association studies provide an unparalleled starting point for the study of schizophrenia aetiology, which may have large impact on treatment in the future. Nevertheless, earlier epidemiological, pathological, neuroimaging and neurochemical studies have been fundamental in exploring schizophrenia aetiology, and some of the landmark studies will be discussed in the following sections.

1.3. Neuroanatomy and neuroimaging in schizophrenia

Studies using computed tomography, structural magnetic resonance and diffusion tensor imaging have repeatedly observed brain structure abnormalities in patients with schizophrenia and their first degree relatives, as compared to healthy subjects without family history of psychiatric illness. The most consistent observations are the enlargement of the lateral ventricles (central arrows in Figure 1) (Johnstone et al., 1976; Olabi et al., 2011; Wright et al., 2000), generalised anisotropy of the white matter (Lee et al., 2009; Mitelman et al., 2009; Nazeri et al., 2013; Rametti et al., 2009), white matter abnormalities in the nucleus caudate (Takase et al., 2004), decreased overall brain size, and reduced cortical grey matter volume and weight (right arrows in Figure 1), more prominently in the medial temporal lobe, the superior temporal gyrus, and the insula cortex (Harrison et al., 2003; Honea et al., 2005).

The prefrontal cortex and the hippocampus are from far the most extensively studied brain regions in schizophrenia research, as abnormalities in these areas can explain symptoms due to their association with the limbic system (Bakhshi and Chance, 2015). Functional magnetic resonance imaging (fMRI) studies have indeed reported abnormal activation of the prefrontal cortex in schizophrenia subjects during working memory, emotion processing, word fluency and decision-making tests (Gur and Gur, 2010; Wong and Van Tol, 2003), as well as a generally increased activation of the hippocampal CA1 region in patients (Talati et al., 2014). Such reports of neuroimaging findings in association with disease are frequent (Adriano et al., 2012; Arnsten, 2013; Nelson et al., 1998; Zhou et al., 2015), but are not of diagnostic value as there is considerable overlap between healthy and affected individuals.

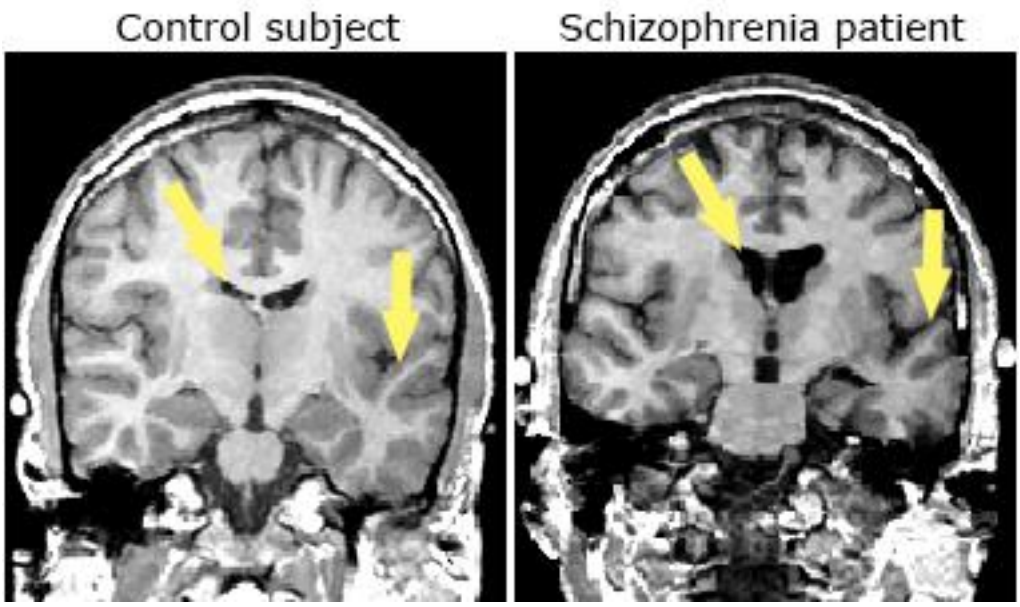


Figure 1. Coronal view of a control subject and a schizophrenia patient using magnetic resonance imaging (MRI). The central yellow arrows demonstrate the increased lateral ventricles in a chronic schizophrenia patient. The right-hand side arrows point to an area of thin cortex, a feature also associated with schizophrenia. Images obtained from the Harvard University Schizophrenia Project.

There is currently intensive research being undertaken in the development of integrative and more sophisticated approaches for analysing neuroimaging data. These include the categorisation of subjects based on genetic data (Dima and Breen, 2015; Dudbridge, 2013; Oertel-Knochel et al., 2015), the coupling of different scan methods by means of multimodal neuroimaging (Sui et al., 2015), and the use of machine learning algorithms for the detection of subtle changes in morphology or function (Sun et al., 2009; Yang et al., 2010). While a pathognomonic signature for schizophrenia is yet to be found (Moncrieff and Middleton, 2015), the development of more accurate analysis methods and more powerful scanners collectively hold a significant potential for neuroimaging as an emerging tool for diagnosis and for investigation of disease aetiology.

1.4. Pathophysiology of schizophrenia

Markers of neurodegeneration are not distinctively observed in the schizophrenia brain (Weinberger, 1987). The anatomic deviations associated with schizophrenia are consequently not likely to be a product of neurodegeneration, and seem more consistent with issues in connectivity and plasticity (Cao et al., 2016). The observed macroscopic abnormalities in neuroimaging and pathological studies have been linked with cytoarchitectural changes, which include reduced number of dendritic spines (Figure 2), neurites, cell soma size and neuropil (Glantz and Lewis, 2000; Lewis and Gonzalez-Burgos, 2008; Selemon and Goldman-Rakic, 1999; Walker et al., 2002), which collectively increase cortical density (Selemon et al., 1995). These data suggest abnormal neuronal connectivity in schizophrenia, supported by recent studies with human induced pluripotent stem cells (hiPSC), in which hiPSC-derived neurons from schizophrenia patients were found to develop fewer synapses and neurites, as well as to display decreased levels of postsynaptic density protein 95 (PSD95) and glutamatergic receptors (Brennand et al., 2011) when compared to neurons derived from healthy controls. Abnormal inhibition of the frontal cortex by GABAergic interneurons has also been hypothesised to play a role in schizophrenia (Nakazawa et al., 2012).

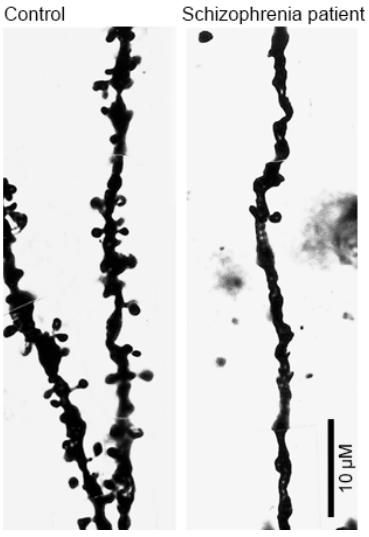


Figure 2. Neurites from the prefrontal cortex of a healthy control and a patient. Stained with Golgi's method and analysed under bright field microscopy. Adapted from Glantz and Lewis (2000). A decreased number of dendritic spines has been detected in schizophrenia, hypothesised to account for the cortical thinning observed in patients, more recently replicated elsewhere (Konopaske et al., 2014).

The glutamatergic system, too, has been implicated in this disease, with transgenic rodents selectively lacking N-methyl-d-aspartic acid-type (NMDA) receptors in cortical and hippocampal interneurons displaying a vast range of molecular and cognitive characteristics that resemble schizophrenia (Belforte et al., 2010). Additionally, oxidative stress has been found to be a plausible explanation for the establishment of schizophrenia symptoms, and preliminary clinical trials with N-acetyl cysteine found that this antioxidant agent was associated with a significant improvement of cognitive symptoms in patients (Berk et al., 2008; Bitanihirwe and Woo, 2011). Another non-mutually exclusive hypothesis regards the role of neuroinflammation in schizophrenia, as *post-mortem* studies found elevated markers of neuroinflammation in the frontal and temporal cortices of patients (Fillman et al., 2013; Radewicz et al., 2000). Considering that maternal infection is an apparent risk factor for schizophrenia (Khandaker et al., 2013), it is hypothesised that early-life exposure to certain pathogens has a long lasting local activation in the microglia caused by circulating pro-inflammatory cytokines, which have been additionally observed to alter early neural architecture (Monji et al., 2013). These inflammatory molecules are linked to GABAergic hypofunction in certain brain areas, which can subsequently affect glutamatergic and dopaminergic signalling. Parallel studies provided evidence to support the role of inflammation and synaptic connectivity in the pathophysiology of schizophrenia, such as the global gene expression profiling of *post-mortem* brains in case-control design studies, which implicate components of the immune system and of formation and myelination of synapses in schizophrenia (Hakak et al., 2001; Mirnics et al., 2000; Mirnics et al., 2006; Roussos et al., 2012). An enrichment for genes involved in the immune system and synaptic function has also been reported at loci implicated by genome-wide association studies of schizophrenia (Lips et al., 2012). Collectively, these data are consistent with abnormal connectivity, signalling and plasticity, as well as a possible immune component in schizophrenia (Bakhshi and Chance, 2015; Harrison and Weinberger, 2005), corroborating the hypothesis that schizophrenia is a heterogeneous condition.

1.5. Neurochemistry of schizophrenia

Dopamine is likely the most extensively studied neurotransmitter in schizophrenia, mostly due to early neurochemical findings (Reynolds, 2005). In the 1950s, administration of derivatives of typical antipsychotics (phenothiazines) as tranquilisers led to the serendipitous development of the first antipsychotic, chlorpromazine, in that decade (Ramachandraiah et al., 2009). These drugs were found to act as dopamine receptor antagonists, with their capacity to ameliorate psychotic symptoms directly correlating with their binding to dopamine receptors (Seeman and Lee, 1975). It was also found that amphetamine, a strong stimulator of the dopaminergic system (an agonist of the trace amine-associated receptor 1, TAAR1), mimics positive symptoms in healthy individuals, and exacerbates these symptoms in patients (Carlsson 1978). Together, these observations led to the 'dopamine hypothesis of schizophrenia' (Carlsson, 1978; van Rossum, 1966), which postulated that schizophrenia (or at least the positive symptoms of the disorder) arises from overactive dopamine function. These findings have been more recently supported by functional imaging studies, which have found hyperactive subcortical dopaminergic pathways in patients (Laruelle, 2000).

The dopamine hypothesis alone, however, cannot explain all features of schizophrenia, such as the negative and cognitive symptoms. Pharmacological evidence for the implication of other neurotransmitters in this psychiatric disease soon came to place. The administration of NMDA receptor antagonists such as phencyclidine (PCP) and ketamine were found to temporarily trigger positive and negative symptoms in healthy individuals by altering glutamatergic transmission (Krystal et al., 1994; Lahti et al., 2001; Luisada, 1978). In a small clinical trial, sarcosine has been found to significantly improve negative symptoms in patients (Hashimoto, 2010). Sarcosine is known to inhibit glycine uptake, causing increased glycine bioavailability in the synaptic cleft, therefore allowing this amino acid to act as co-agonist to glutamate, modulating NMDA receptor activity. Gamma-aminobutyric acid (GABA) was additionally implicated in schizophrenia pathology,

since it is intimately associated with glutamate transmission, although it is generally associated with inhibitory action to glutamatergic excitatory neurons. Decreased concentration of cortical GABA in patients has been observed (Perry et al., 1979), as well as decreased cortical glutamic acid carboxylase, a rate-limiting enzyme in the synthesis of this neurotransmitter (Bird, 1985; Curley et al., 2011). As mentioned in the previous section, decreased markers of GABAergic interneurons in the cortex and in the hippocampus have additionally been found in patients, supporting a role for altered GABA signalling in schizophrenia (Nakazawa et al., 2012; Reynolds, 2005). The role of serotonin in schizophrenia is based on evidence of abnormal levels of this neurotransmitter in blood and cerebrospinal fluid of patients (Bleich et al., 1988). Early *post-mortem* studies confirming increased dopamine receptors in the caudate nucleus of patients also showed reduced levels of serotonin receptors in the prefrontal cortex (Mita et al., 1986). Pharmacological studies with lysergic acid diethylamide (LSD), which binds to several G-coupled receptors including serotonin receptors, suggested a role for serotonin in schizophrenia, since this drug is known to cause hallucinations in healthy subjects (Nichols, 2004). These findings support the involvement of several neurotransmitter systems in schizophrenia neurobiology.

1.6. Epigenetics of schizophrenia

Epigenetic processes are essential for neural development and function. The epigenome describes the chemical modifications encoded 'on top' of the genetic material (hence the Greek prefix '*epi*'), which may be inherited or modified by environmental factors (Shorter and Miller, 2015). These chemical modifications alter the chromatin structure and therefore may modify access of transcription factors and polymerases to the DNA coiled in the nucleosome (the structural unit of the chromosome in eukaryotes). Chemical modifications may occur directly at the DNA level, for example with the addition of methyl or hydroxymethyl groups to cytosine bases (Kriaucionis and Heintz, 2009) or, alternatively, at the DNA-coiling proteins histones in the nucleosome, which can undergo the addition of methyl, acetyl, ubiquityl groups, etc. (Figure 3) (Tan et al., 2011).

DNA methylation is the most extensively studied epigenetic process, and is typically associated with transcriptional repression. DNA methylation differences in schizophrenia have been reported for a number of candidate genes, including *BDNF* (brain-derived neurotrophic factor), which encodes for a protein that supports the growth and differentiation of neurons (Wong et al., 2010); *GAD1 (*glutamic acid decarboxylase 67), which encodes the enzyme responsible for GABA synthesis (Huang and Akbarian, 2007); and *RELN* (reelin), which encodes a protein involved in cell-cell interaction and neuronal migration during brain development (Grayson et al., 2005). In the first genome-wide study of DNA methylation in the major psychoses (schizophrenia and bipolar disorder), Mill and colleagues reported widespread changes in DNA methylation in the frontal cortex between patients and controls, including loci involved in glutamatergic, GABAergic function and brain development (Mill et al., 2008). Genome-wide studies have identified numerous genetic variants that are associated with DNA methylation (methylation quantitative trait loci; mQTL) in both the adult and developing human brain (Gibbs et al., 2010; Hannon et al., 2016). Moreover, Hannon and colleagues found that risk variants identified in a large-scale GWAS of schizophrenia (Schizophrenia Working Group of the PGC, 2014) were enriched for foetal brain mQTL, supporting a neurodevelopmental component to the disorder (Hannon et al., 2016). A role for acetylation of histones has also been implicated in schizophrenia, as expression of histone deacetylase 1 was found increased in the cortex of patients (Sharma et al., 2008). In addition, the Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015) found an enrichment for genes encoding histone methylases in genome-wide association signals from schizophrenia, depression and bipolar disorder cohorts combined.

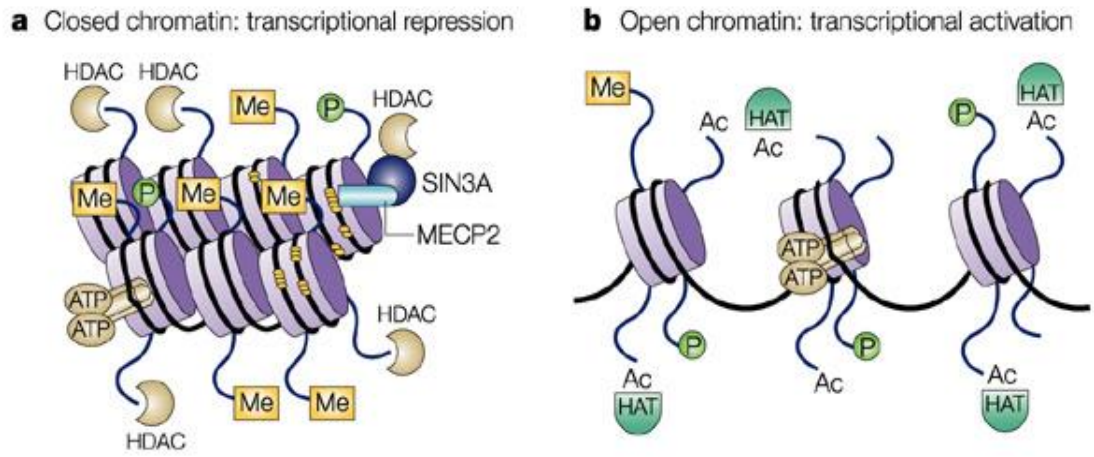


Figure 3. Representation of two extremes of chromatin status in the nucleus, as dictated by epigenetic modifications. (**a**, open; **b**, closed) The DNA molecule (black line) is wrapped around histone octamers (purple). Methylation of cytosine residues (yellow dots) contribute to the chromatin state, and so do other chemical modifications on the histones, such as methylation (Me, yellow), phosphorylation (P, green) or acetylation (Ac). Histone deacetylases (HDACs, yellow) and histone acetyltransferases (HATs, green) are enzymes that modify epigenetic marks in histones. Proteins such as methyl-CpG-binding protein (MECP2, blue) can target DNA for deacetylation. Figure extracted from the review of Johnstone (2002).

1.7. Environmental risk factors for schizophrenia

Several pre- and postnatal chemical, biological and psychosocial stressors are known to increase risk for schizophrenia. Maternal infection (Blomström et al., 2016), malnutrition (Kirkbride et al., 2012), stress (Khashan et al., 2008) and complications at birth leading to hypoxia (Hultman et al., 1999) increase the likelihood of the offspring to develop schizophrenia. Infection with pathogens such as *Toxoplasma gondii* (Torrey et al., 2012), or exposure to Cannabis (Radhakrishnan et al., 2014), migration (Cooper, 2005) and urbanicity later in life (Krabbendam and van Os, 2005), as well as childhood adversity (Varese et al., 2012), have also been reported as risk factors for the disorder. The fact that a genetically identical sibling of a schizophrenia patient has approximately 50% risk of developing this disorder emphasises a role for non-genetic factors in risk (Gottesman, 1991).

Although the biological processes driving these associations with schizophrenia remain unclear, some mechanisms have been proposed. For example, brain sections from rodents infected with *T. gondii* showed intense staining for dopamine, and this was associated with the expression of the parasite's own tyrosine hydroxylase (enzyme that catalyses the tyrosine conversion to dopamine) in the host tissue, providing a rationale for the behavioural changes induced by the pathogen (Prandovszky et al., 2011). It is plausible that the host's immune response is also involved in this process, as infection of rodents with *T. gondii* caused the release of interleukins (ILs) 12 and 18, interferon-gamma (IF-gamma) and tumour necrosis factor-alpha (TNF-alpha) (Silva et al., 2002). Some pro-inflammatory interleukins were also found to be increased in the serum of schizophrenia individuals, as compared to control subjects (Cazzullo et al., 1998; Maes et al., 1995). These studies, collectively, provide evidence for the role of the immune system and neuroinflammation in schizophrenia. Pathway analysis of the genetic loci implicated by GWAS in depression, schizophrenia and bipolar also revealed an enrichment for immune processes such as TGF (transforming growth factor) signalling and activation of B and T cells in these disorders, even after excluding MHC (major histocompatibility complex) signals from the study (Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015).

Finally, the environment can exert its detrimental effects in brain development and function by means of epigenetics, as methylation of DNA is known to play a role in neural plasticity and has been shown to elicit long-lasting effects in gene expression through generations (Levenson et al., 2006; Miller and Sweatt, 2007). It is known, for example, that exposure of infant mice to stressed caretakers cause differential methylation of *BDNF* in the prefrontal cortex, which later on might cause abnormal expression of this important neurotrophic factor (Roth et al., 2009). These findings collectively suggest that there are many factors that contribute to risk for schizophrenia.

1.8. Genetic studies of schizophrenia

Consistent with a large genetic component to schizophrenia, the risk of developing the disorder rises exponentially with the level of genetic relatedness to an affected individual (Gottesman, 1991). Thus, schizophrenia risk for a member of the general population jumps from 1% to approximately 50% for an individual whose genetically identical sibling suffers from schizophrenia (Figure 4). Moreover, a meta-analysis of twin studies estimated the heritability of schizophrenia to be as high as 81% (Sullivan et al., 2003). Although neurobiological research of schizophrenia has provided broad clues to its pathophysiology, the underlying molecular mechanisms involved in this mental disorder remain unknown. Additionally, neuropathological investigations have traditionally suffered from problems in distinguishing changes that reflect primary disease mechanisms from secondary manifestations of the illness, such as compensatory changes and effects of medication. Genetic studies, in their capacity to identify genes and molecules that play a direct role in susceptibility, are likely to provide valuable insight into schizophrenia aetiology (Bray, 2008; Bray et al., 2010; Harrison and Weinberger, 2005; Insel, 2009).

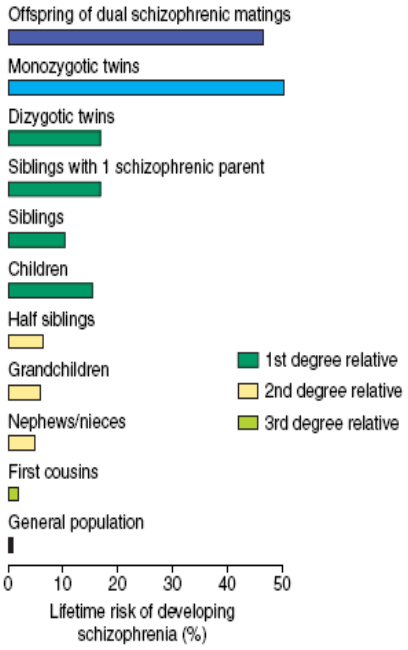


Figure 4. Life-time risk of developing schizophrenia according to familial relationship to someone with the disorder. Compared with the general population risk of 1%, risk is approximately 50% where a monozygotic twin is affected. Extracted from Bray and Owen (2001).

1.8.1. Early findings

Early molecular investigative tools to elucidate the role of genetic variants predisposing to schizophrenia included genetic linkage studies and positional or functional candidate gene association studies. Linkage studies provide a genome-wide scan for susceptibility loci by investigating the co-segregation of chromosomal markers with the illness within families. This approach is able to identify chromosomal regions harbouring variants that make a large contribution to risk (e.g. the Huntington's disease gene) (Gusella et al., 1983), but has showed limited success in identifying schizophrenia susceptibility loci (Sullivan, 2008). Genetic association studies, which usually compare allele or genotype frequencies between groups of affected 'cases' and unaffected 'controls', potentially allow identification of risk variants of weaker effect. 'Functional candidate gene' association studies of schizophrenia included analyses of variants within genes involved in neurotransmission and other neural functions thought to be disturbed in schizophrenia (e.g. *DRD2, COMT, 5HT2A*), but results were largely inconclusive (Sullivan, 2008). 'Positional candidate gene' association methods investigate association within regions implicated by linkage studies. These studies implicated genetic variation in several genes to schizophrenia susceptibility (e.g. *DTNBP1* on chromosome 6p22 (Straub et al., 2002); *NRG1* on chromosome 8p12 (Stefansson et al., 2002)), but none conclusively. More radical, 'structural' genetic variations have also long been implicated in schizophrenia risk, such as a chromosomal translocation between chromosomes 1q42 (affecting the gene disrupted in schizophrenia 1*, DISC1*) and 11q14, which co-segregates with schizophrenia and other mental illnesses in a large Scottish family (Millar et al., 2000), and large deletions on chromosome 22q11.2, which cause velocardiofacial syndrome (Murphy et al., 1999). Nonetheless, these rare risk variants are rare even in cohorts of schizophrenia patients (<1%) and therefore they do not represent the biological causes of this disorder in their entirety (Bassett and Chow, 2008).

These early findings provided the first clues about the extremely polygenic and multifactorial nature of schizophrenia. It was only with the introduction of next-generation sequencing (NGS), genome-wide association studies (GWAS) and the formation of large collaborative consortia that research into schizophrenia could identify numerous rare and common susceptibility variants, which are now known to collectively contribute to risk for disease (Fromer et al., 2014; Purcell et al., 2014; Ripke et al., 2013; Schizophrenia Working Group of the PGC, 2014).

1.8.2. Common and rare variants in schizophrenia

Genetic variation is often categorised into 'rare' and 'common' genetic variants, with rare alleles usually defined as those with population frequencies less than 1%. The association of rare and common variants to schizophrenia has been observed to overlap to some extent (Richards et al., 2016), suggesting that they may increase susceptibility through shared downstream pathways. These variants have also been observed to contribute to genetic susceptibility independently, which supports a polygenic risk model (Tansey et al., 2016). The identification of these classes of genetic variation in association with schizophrenia requires different methodological approaches. As discussed above, rare alleles are generally rare even within groups of affected subjects, but can be associated with large effect sizes (Vassos et al., 2010). The methods of choice for identifying rare variants currently include whole-exome/whole-genome sequencing studies and DNA microarrays. Although all these methods are considered genome-wide, microarrays limit the study to known variants, whereas sequencing is more expensive but allows the identification of novel variants. Rare variants can be subdivided into copy number variants (CNVs), defined as large genetic deletions or duplications frequently involving multiple genes, and rare single nucleotide variants (SNVs) associated with non-synonymous mutations that either impact protein structure or cause premature stop codons. The latter can be exemplified by the rare SNVs which cause loss of function in the schizophrenia susceptibility gene *SETD1A* (SET domain containing 1A)

(Singh et al., 2016), while the former includes CNVs at chromosomes 16p12.1, 1p36.33, 1q21.1, 15q13.3 and 22q11.2, amongst others (Lichtenstein et al., 2009; Rees et al., 2014; Takata et al., 2013). Rare variants were found to collectively impact on components involved in neural plasticity, such as members of the postsynaptic density proteins, NMDA receptors, and the cytoskeleton, supporting their role in schizophrenia pathology (Hall et al., 2015).

Common alleles, on the other hand, are associated with small effects on schizophrenia risk, with odds ratios typically < 1.2 (Bray and Hill, 2016). The method of choice for detecting common variants associated with a trait is through genome-wide association studies (GWAS). In GWAS, the allelic frequencies from common polymorphisms identified through the International HapMap Consortium (2003) and the 1000 Genomes Project Consortium (2012) are tested for association with a trait using SNP (single nucleotide polymorphism) microarray chips. Next-generation sequencing, however, will likely replace microarray technology in these studies (Sharma et al., 2014). In any case, because of multiple testing and small individual effects contributing to complex traits, large-scale cohorts sometimes involving hundreds of thousands of individuals are required to provide statistical power for this sort of analysis. Moreover, as for most GWAS results, a large proportion of genome-wide significant association signals, approximately 90% of the current 108 genetic loci implicated in schizophrenia, is non-coding and therefore hypothesised to impact on gene expression, producing effects that are not obvious to the protein structure (Schizophrenia Working Group of the PGC, 2014). Although common variants are individually associated with small effects in risk, they collectively account for a third to half of the genetic burden involved in schizophrenia (International Schizophrenia Consortium et al., 2009; Ripke et al., 2013).

In summary, schizophrenia is a highly heritable, polygenic and multifactorial condition. A range of common and rare variants are hypothesised to act in concert with pre- and postnatal biological, chemical and psychosocial stressors to dictate risk for this mental disorder, which can be represented in a threshold model (Figure 5). It is likely that many aspects of previous hypotheses of schizophrenia, including the dopaminergic, the glutamatergic, the GABAergic, and the neurodevelopmental hypotheses, play a significant role in schizophrenia aetiology, either in subgroups of patients or in particular time points during development. But only molecular studies will be able to confidently identify the abnormal processes ultimately driving symptoms. Finally, there is a considerable genetic overlap between schizophrenia and other psychiatric diseases, particularly bipolar disorder and major depression (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Georgieva et al., 2014; International Schizophrenia Consortium et al., 2009; Lee et al., 2013). Thousands of genes and non-coding loci are estimated to contribute to risk (International Schizophrenia Consortium et al., 2009; Kiezun et al., 2012; Ripke et al., 2013), but molecular mechanisms underpinning such associations remain unclear for all these disorders.

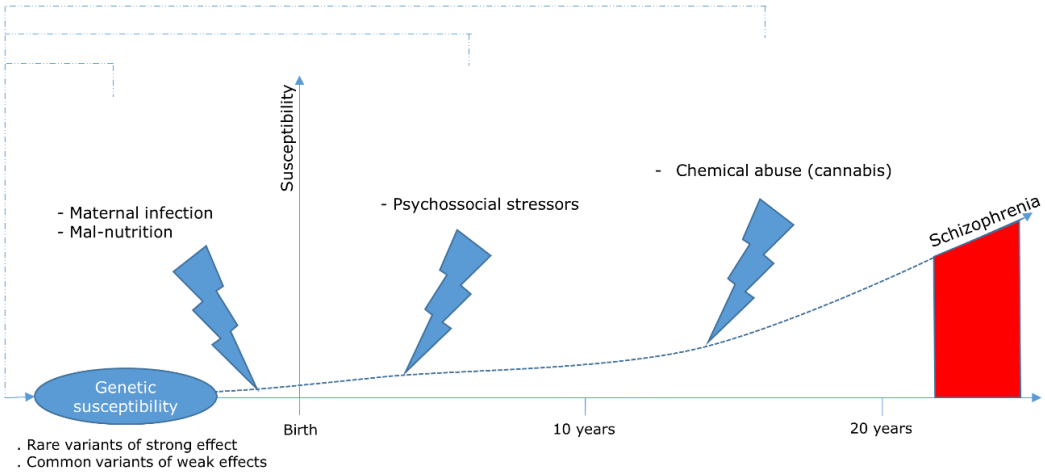


Figure 5. Representation of the threshold model of schizophrenia. Genetic susceptibility is dictated by rare and common variants, which interact with pre- and postnatal biological, chemical and psychosocial stressors in a multivariate threshold model of risk. It is plausible that genetic variation is also associated with individual response to different stressors, therefore contributing indirectly towards risk for disease.

1.8.3. Genome-wide association studies (GWAS) of schizophrenia and follow-up studies

Genome-wide association studies are known as 'hypothesis-free' analyses because they test for association between a trait and common variants across the genome, rather than focusing on specific candidate genes. GWAS can investigate thousands of subjects in case-control design studies (such as for most diseases), or quantitative trait design studies (such as for characteristics like height or body mass index). Currently, SNP chips are used to investigate the status of approximately one million common genetic variants, for which genotype is further imputed to approximately 10 million common variants that contribute to individual differences. Comparing frequencies from so many variants necessitates multiple testing correction, for instance Bonferroni, which reflects the *P*-value of 0.05 corrected for a million independent tests (i.e. $P < 5x10^{-8}$). Common variants are usually of small effect sizes, and therefore reaching this threshold of 'genome-wide significance' requires the analysis of large cohorts (McCarthy et al., 2008).

The largest and most recent GWAS of schizophrenia, performed by the Psychiatric Genomics Consortium (PGC), analysed 150,064 individuals and implicated 128 common variants distributed in 108 genomic loci in susceptibility to schizophrenia[1] (Figure 6) (Schizophrenia Working Group of the PGC, 2014). Although the Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015) found an enrichment for biological processes involved in neuronal signalling, immune system and histone methylation in GWAS signals of schizophrenia, major depression and bipolar disorder combined, the contribution of individual signals to risk remain poorly characterised. It is likely that understanding their biological meaning will provide new opportunities for treatment.

[1] This study was performed by the Psychiatric Genomics Consortium (PGC), for which the Schizophrenia Workgroup itself includes over 400 investigators from 40 countries.
Source: www.med.unc.edu/pgc/pgc-workgroups. Accessed on 20th Sep 2016.

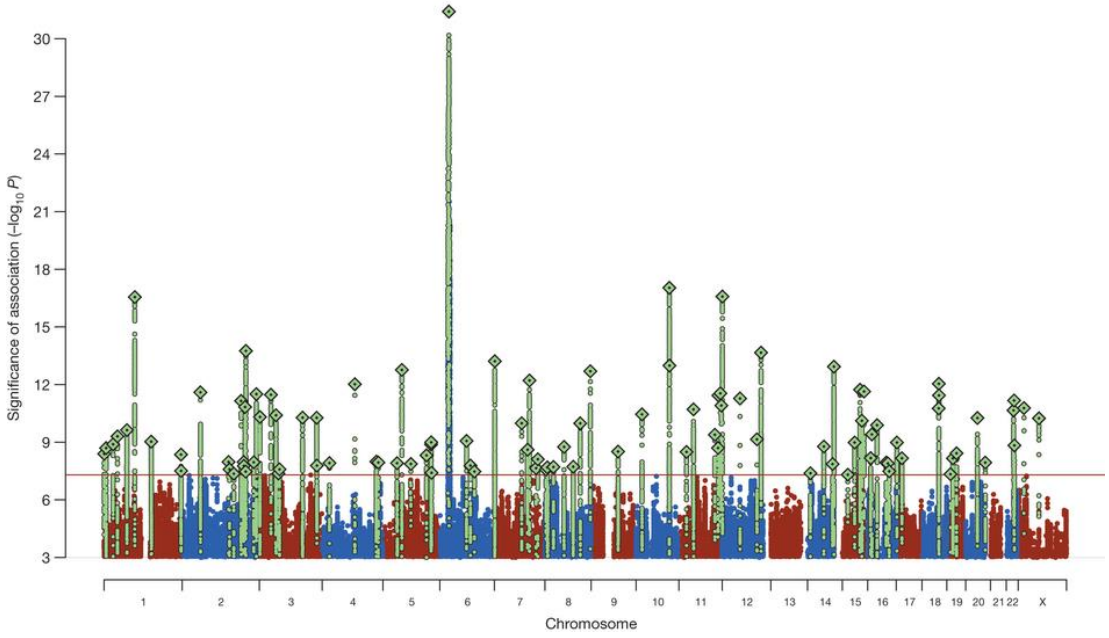


Figure 6. Regional association plot of the latest schizophrenia GWAS results, as presented by a Manhattan plot. The X-axis indicates genomic coordinates ranging from chromosome 1 to X. The Y-axis indicates the confidence of association by means of -$\log_{10}$P. The highest peak of association corresponds to the MHC locus on chromosome 6, followed by a hit on chromosome 1 close to *DPYD* and *MIR137*, followed by an association signal on chromosome 10q24, main subject of this thesis. Figure extracted from the article "Biological insights from 108 schizophrenia-associated genetic loci" (Schizophrenia Working Group of the PGC, 2014).

One of the obstacles in interpreting individual GWAS signals is due to a phenomenon called linkage disequilibrium (LD), in which genotype at neighbouring variants correlate with each other within a population. This causes a significant proportion of the association signals to span large genomic regions (Bray and Hill, 2016), with a notable example of this being the schizophrenia association signal on chromosome 6, at the MHC locus (Figure 6) (Schizophrenia Working Group of the PGC, 2014). This region has a complex genetic structure, containing 18 highly polymorphic human leukocyte antigen (*HLA*) genes, with linkage disequilibrium spanning over ~8 million bases over hundreds of immune and non-immune genes. In order to determine the true functional variants driving association with schizophrenia at this locus, Sekar et al. (2016) combined RNA-sequencing (RNAseq-) data with a range of sophisticated molecular tools and bioinformatic analyses to identify the underlying risk mechanisms. Association of complement 4 (C4) expression with genetic structure information was performed using digital PCR and

RNA-seq data from adult *post-mortem* brains. Expression of C4 was observed elevated in schizophrenia as compared to controls in several brain regions, and its knockdown in mice was found to cause deficits in synaptic pruning, a process hypothesised to be aberrant in schizophrenia (Feinberg, 1982). Sophisticated computational analyses of RNA-seq data revealed that the schizophrenia hit on the MHC locus is tagging at least three independent association signals at the locus, for which only one is responsible for driving expression of C4.

Furthermore, approximately 90% of genome-wide association signals for schizophrenia are non-coding (Schizophrenia Working Group of the PGC, 2014), and therefore it is not obvious which genes are affected and how they are functionally altered (Barr and Misener, 2016). *Cis*-regulatory effects, in which functional risk variants alter the expression or splicing of transcripts encoded by the same chromosome strand on which they reside, have been studied using a variety of methods. These can be broadly divided into those which measure a gene's overall expression and correlate this with genotype, and those which measure and compare gene expression arising from each parental chromosome (allelic expression). The former approach typically combines microarray or RNA-Seq measures of gene expression with genome-wide genotyping (GTEx Consortium, 2015; Myers et al., 2008) to identify expression quantitative trait loci (eQTL), arbitrarily denoting those that are close to the gene (e.g. within 1Mb) as 'cis-acting' and more distant eQTL as 'trans-acting'. In the latter approach, allele-specific expression measures can be assessed on individual genes using qPCR (Chen et al., 2008; Udler et al., 2007) or single nucleotide extension (SNaPshot) (Bray et al., 2003a; Bray et al., 2003b) methods, or can be performed genome-wide utilising SNP arrays (Lo et al., 2003; Serre et al., 2008) or RNA-Seq (GTEx Consortium, 2015; Heap et al., 2010), specifically detecting *cis*-regulatory variation.

The detection of *cis*-regulatory effects using SNaPshot primer extension, as will be further discussed in the next chapter, is achieved by coupling measures of allele-specific expression using single-nucleotide extension with genotypic

information at risk variants of interest. This method uses an expressed 'reporter' SNP in genes of interest, for testing allelic expression in individuals that are heterozygous for that variant, allowing the discrimination between the RNA molecules transcribed from each chromosomal copy. It has been previously used to assess potential *cis*-regulatory effects of risk genotypes on a number of candidate genes for schizophrenia (Bray et al., 2003a). One of the advantages of using allele-specific measures, as opposed to methods based on a gene's overall expression, is that it allows the accurate measurement of *cis*-regulatory effects while controling for *trans*-acting influences and environmental variables on gene expression, as these confounders will usually act upon both alleles to the same extent.

Cis-regulatory mechanisms can be also specific to particular tissues or cell types, at precise time points. This occurs because functional risk variants enhancing or repressing transcription are most likely binding sites for transcription factors, for which their access is additionally subject to chromatin status (closed or open conformation; Figure 3), as dictated by DNA sequence and epigenetic markers. Nonetheless, time- and tissue-dependent *cis*-regulatory effects can be investigated by testing multiple samples that consist of different developmental time points, or different brain regions. As example, Hill and Bray (2012) observed reduced expression of *ZNF804A* in association with risk variants for psychosis in the second trimester foetal brain, but not in first trimester foetal brain or in the adult dorsolateral prefrontal cortex, substantia nigra or hippocampus. Alternatively, description of the signal can be achieved using single-cell RNA-sequencing or tissue microdissection.

Additionally hindering the identification of causal susceptibility mechanisms, functional variants may impact specific transcripts of a gene, potentially novel, non-annotated transcript variants (Bray and Hill, 2016). Genes typically encode for multiple transcripts, which may collectively exhibit different decay properties or target specific subcellular domains, or potentially produce different protein isoforms (Jung et al., 2012). In the case of *ZNF804A*, for example, Tao et al. (2014) found that the putative psychosis-related risk transcript of *ZNF804A, ZNF804A^{E3E4},* encodes

a novel protein of unknown function, which lacks the zinc-finger domain. In order to detect this novel transcript, Tao and collaborators used RNA-seq to guide the design of specific primers for rapid amplification of cDNA ends (RACE). Finally, the authors found in *post-mortem* tissue that this susceptibility transcript was less expressed in schizophrenia patients when compared to controls.

Once a gene is implicated in risk, its product or expression levels can be investigated in different developmental points, brain areas or tissues by using publicly available databases, such as the Protein Atlas (www.proteinatlas.org), the Allen Brain Atlas (www.brain-map.org) and the Human Brain Transcriptome Atlas (hbatlas.org). Additionally, risk mechanisms can be simulated *in vitro* once their effect on the genetic target is identified (e.g. up- or downregulation). This can be pursued by specifically manipulating expression of the susceptibility genes by, for example, silencing them using RNA interference (RNAi), or overexpressing them by transfecting expression constructs into cellular models of interest. Consequently, risk mechanisms can be emulated in cellular models so that they may serve as drug screening platforms. Alternatively, transcriptomic changes associated with the risk mechanism may be analysed by microarrays or RNA-seq in order to identify biological pathways affected. The knockdown of *ZNF804A* full length, for example, was initially reproduced in neural progenitor cells (NPCs) to simulate the putative risk mechanism involving its decreased expression in schizophrenia during foetal development, which found genetic pathways involved in cell adhesion to be altered (Hill et al., 2012). In this context, the aim of this thesis is to functionally characterise one of the most robust genome-wide association signals of schizophrenia to date, located on chromosome 10q24. Some of these methods will be further discussed during the following chapters.

1.8.3.1. Chromosome 10q24

Robustly associated with schizophrenia in previous studies (Aberg et al., 2013; Guan et al., 2016; Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011), the chromosome 10q24 locus harbours the third most significant association signal to emerge from the largest GWAS of schizophrenia to date (Schizophrenia Working Group of the PGC, 2014). This locus is also genome-wide significant for the five disorders of the Psychiatric Genomics Consortium combined (major depression, bipolar disorder, schizophrenia, attention deficit and hyperactivity disorder, and autism) (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013), suggesting that variation at this locus may contribute to risk to psychiatric diseases in general. As for most GWAS hits, the most significantly associated variants are non-coding (and therefore expected to impact on gene expression), and located in a region of strong LD that spans approximately 400 kb, hindering the identification of the true susceptibility gene(s) (Figure 7). Risk SNPs at this locus (or variants in LD with them) have been associated with social cognition and morphometric brain differences implicated in schizophrenia (Ohi et al., 2013; Ohi, 2015; Rose et al., 2014). However, the molecular mechanisms underpinning these events and their association with schizophrenia remain unknown. The focal positional candidates encompassed by strong linkage disequilibrium in the region include *BORCS7* (BLOC-1 related complex subunit 7), *AS3MT* (arsenite methyltransferase), *CNNM2* (cyclin M2) and *NT5C2* (cytosolic 5'-nucleotidase II). None of these genes have a clear established role in neural function, but their hypothetical roles in the nervous system are described on Table 1.

Table 1. Predicted function of chromosome 10q24 genes in region of strongest LD.

Gene symbol	Name / synonyms	Putative function	Relevance	Furthest taxon found (conservation)
BORCS7	BLOC-1 related complex subunit 7, *C10ORF32*	Lysosomal function and cell migration (Pu et al., 2015)	BLOC-1 has been implicated in neurodevelopment (Ghiani et al., 2010)	Chordates
AS3MT	Arsenic methyltransferase	Metabolism of arsenic compounds (Sumi and Himeno, 2012)	Arsenic poisoning associated with neurological deficits (Tyler and Allan, 2014)	Chordates
CNNM2	Cyclin and CBS domain divalent metal cation transport mediator 2, cyclin 2	Magnesium homeostasis in the kidneys (Stuiver et al., 2011)	Mutations in the gene associated with seizures and abnormal brain development in hypomagnesemia (Arjona et al., 2014)	Eukaryotes
NT5C2	Cytosolic 5'-nucleotidase II, cN-II	Metabolism of intracellular purinergic nucleotides (Itoh, 2013)	Purinergic hypothesis of schizophrenia (Boison et al., 2012; Lara and Souza, 2000)	Eukaryotes

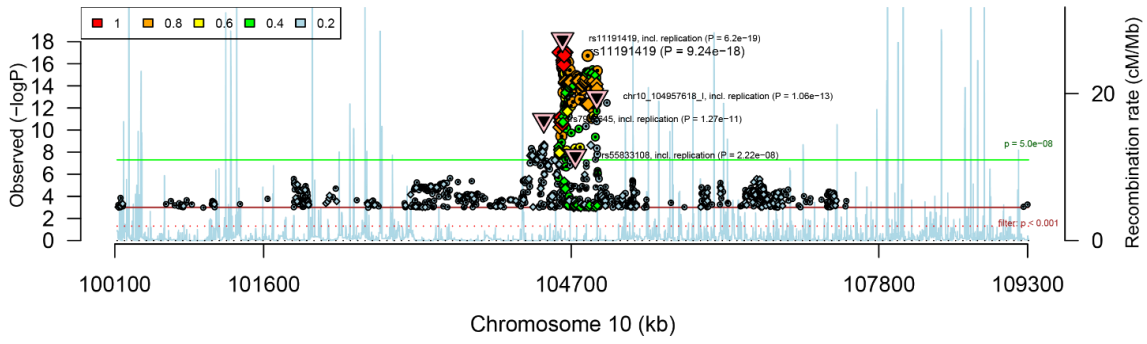


Figure 7. Regional association plot showing the linkage disequilibrium (LD) on chromosome 10 in a 9 Mb-window frame. LD spans an area of approximately 400 kb, on locus q24, comprising of four genes: *BORCS7, AS3MT, CNNM2* and *NT5C2* (see Chapter 2, Figure 1, for a close up view on this region). Variants in linkage disequilibrium with the most significant variants at the locus are represented with the colour scale on the top left panel as function of r^2. SNPs are represented as function of –logP of association with schizophrenia (Y-axis) and chromosomal location (X-axis). The figure was generated using Ricopili (www.broadinstitute.org/mpg/ricopili/).

In the following chapters, a series of methods in molecular biology, biochemistry, bioinformatics and tissue culture will be used to functionally characterise the genome-wide significant association signal on chromosome 10q24 and schizophrenia. The specific aims of this thesis, therefore, are:

- To determine *cis*-regulatory effects on chromosome 10q24 genes associated with genotype at risk variants by means of allele-specific expression measures (Chapter 2);
- To perform a pilot RNA-seq project to characterise the expression of all transcripts from candidate chromosome 10q24 genes in brain tissues implicated in schizophrenia (Chapter 3);
- To investigate expression from the protein encoded by one of the putative schizophrenia susceptibility genes in the adult human brain and in human neural progenitor cells (Chapter 4);
- To reproduce this putative risk mechanism in neural progenitor cells, and to investigate its downstream effects on global gene expression by using microarrays, and applying gene ontology analysis and connectivity mapping in the resulting differentially expressed gene list (Chapter 5);
- To present a summary of these findings, together with a discussion of potential implications and future perspectives for this study (Chapter 6).

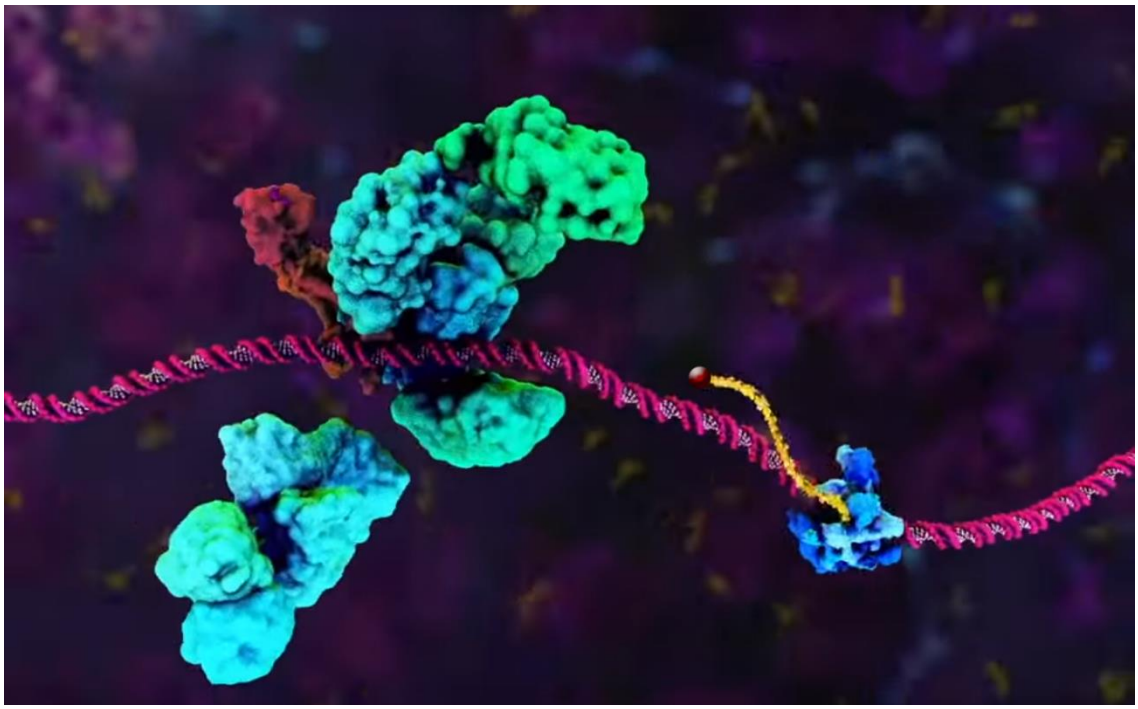


Illustration extracted from the Walter-Eliza Hall, Institute of Medical Research website (video entitled "Sunshine and Vitamin D activation", 00:55min).

Representation of classical gene transcription. RNA (yellow) is transcribed using DNA (pink) as template. Regulatory non-coding sequences in DNA, such as promoters and enhancers, may contain cis-acting variants which alter DNA's affinity to either trans-acting factors (red) or subunits of the RNA Polymerase complex (green and blue). The RNA Polymerase subunit responsible for synthesis is released and reads the information on the DNA molecule in order to produce RNA, while the remaining complex is disassembled.

Chapter 2

Altered *cis*-regulation of *BORCS7*, *AS3MT,* and *NT5C2* in the human brain in association with schizophrenia risk genotypes on chromosome 10q24

2.1. Summary

As sample sizes for genome-wide association studies (GWAS) increase, a growing number of common risk loci associated with schizophrenia have been identified. Most of these signals, however, cannot be immediately translated to susceptibility genes, and therefore the understanding of the risk mechanisms is hindered. This is the case for a region on chromosome 10q24, the third most significant finding from the largest GWAS of schizophrenia published to date, where extensive linkage disequilibrium (LD) and potential long range *cis*-regulatory effects prevent the interpretation of the GWAS signal. This chapter aimed to investigate the effect of genotype at chromosome 10q24 risk variants rs11191419 and chr10_104957618_I on the allelic expression of the positional candidate genes *BORCS7, AS3MT, CNNM2, and NT5C2,* all of which span the region of strong linkage disequilibrium at the locus. Measures of allelic expression were used to assess *cis*-regulatory effects on these genes associated with heterozygosity for risk alleles, in the human foetal brain and in three regions of the adult brain. The results obtained indicate that the risk allele of rs11191419 is associated with a higher expression of the exonic 'reporter' alleles of *BORCS7* and *AS3MT,* which reside on the same chromosomal strand in which the risk alleles are located, while the risk allele of chr10_104957618_I was found to partially counteract these effects. No significant effect of risk genotype was observed on the allelic expression of *CNNM2.* Both risk alleles, however, were found to be concordantly associated with decreased allelic expression of *NT5C2,* as the reporter allele carried on the same chromosome as the risk alleles was found reduced when compared to its allelic counterpart. These data suggest a complex pattern of *cis*-regulatory mechanisms affecting expression of multiple genes on chromosome 10q24, and provide functional evidence to support *BORCS7*, *AS3MT,* and *NT5C2,* as schizophrenia susceptibility genes at the locus.

Note: The results presented in this chapter were first published in the American Journal of Medical Genetics B: Neuropsychiatric Genetics, in Mar 2016, entitled "Genome-wide significant schizophrenia risk variation on chromosome 10q24 is associated with altered *cis*-regulation of *BORCS7, AS3MT,* and *NT5C2,* in the human brain" (Duarte et al., 2016) (Appendix 1).

2.2. Introduction

Amongst the first loci to emerge from large genome-wide association studies of schizophrenia, chromosome 10q24 is one of the most extensively replicated, and it harbours the third top genome-wide significant association signal in the latest study (Aberg et al., 2013; Guan et al., 2016; Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Schizophrenia Working Group of the PGC, 2014). Variation at the locus is additionally genome-wide significant for the five disorders of the Psychiatric Genomics Consortium combined (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013), suggesting that it confers risk to psychiatric illness in general.

The top risk variants at the locus are non-coding and located in a region of high linkage disequilibrium, hindering the identification of the actual susceptibility genes at the locus, and thus limiting its value for improving biological understanding of the associated disorders. Expression studies are therefore required to investigate the regulatory mechanism driving the genetic association signal. In this context, expressed single nucleotide polymorphisms (SNPs) can be used as 'reporter' SNPs to distinguish and relatively quantify RNA transcribed from each parental chromosome in samples of individuals that are heterozygous for the expressed polymorphism. The detection of the *cis*-regulatory effects associated with risk alleles is possible by correlating genotype at risk variants with measures of allelic expression in candidate genes (Bray et al., 2003a). As opposed to classic measures of gene expression like microarray and RT-qPCR, relative estimation of allele specific expression inherently controls for confounding effects such as RNA integrity, age of tissue and *post-mortem* delay, as these events would likely act upon both alleles to the same extent.

In this study, the positional candidate genes in the region of strong linkage disequilibrium on chromosome 10q24 (Figure 1) were screened for potential *cis*-regulatory effects elicited by the risk variants. These genes include *BORCS7* (BLOC-1 related complex subunit 7), *AS3MT* (arsenite methyltransferase), *CNNM2* (cyclin M2) and *NT5C2* (cytosolic 5'-nucleotidase II). Measures of allelic expression were used in *post-mortem* tissue from second trimester foetal brains and three adult brain regions implicated in schizophrenia pathology (the dorsolateral prefrontal cortex, the hippocampus and the nucleus caudate) (Arnold et al., 2015; Callicott et al., 2000; Fillman et al., 2013; Galindo et al., 2016; Honea et al., 2005; Mita et al., 1986; Nelson et al., 1998; Takase et al., 2004), since regulatory mechanisms can be confined to specific brain areas (Buonocore et al., 2010; Gibbs et al., 2010; Ramasamy et al., 2014) or developmental stages (Hill and Bray, 2012; Tao et al., 2014).

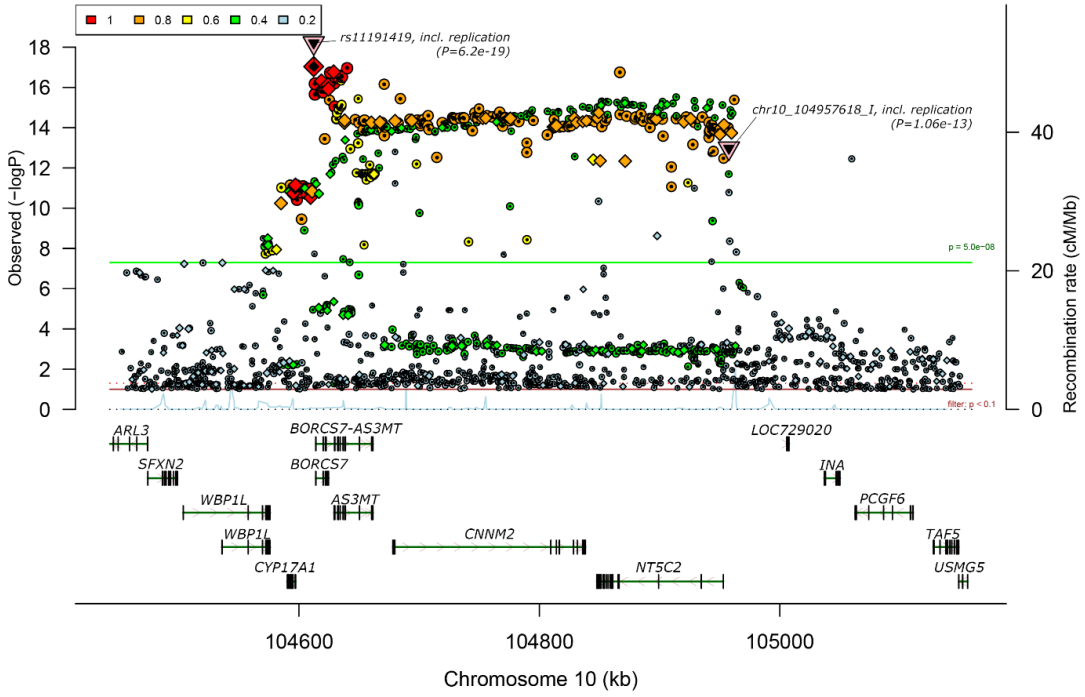


Figure 1. Regional association plot showing high linkage disequilibrium on chromosome 10q24. The strongest variants in association with schizophrenia are pinpointed (rs11191419 and chr10_104957618_I) as per the latest GWAS (Schizophrenia Working Group of the PGC, 2014). Variants in linkage disequilibrium with the most significantly associated variant are represented with the colour scale on the top left panel as function of r^2. SNPs are represented by as function of −log P value of association with schizophrenia (Y-axis) by chromosomal location (X-axis). The figure was principally generated using Ricopili (https://www.broadinstitute.org/mpg/ricopili/) and taken from Duarte et al. (2016).

2.3. Methods

2.3.1. Brain samples

Post-mortem human brain samples from 116 unrelated adults were obtained from the London Neurodegenerative Diseases Brain Bank (UK). All subjects were free from psychiatric or neurological diagnosis at the time of death. Whole brain samples from 95 second trimester human foetuses (13-23 post-conception weeks) resulting from elective abortions were provided by the MRC Wellcome Trust Human Developmental Biology Resource (HDBR, University College London, UK). The mother's consent for research, including genetic analysis, was obtained by the HDBR. Ethical approval was provided by The Joint South London and Maudsley and The Institute of Psychiatry NHS Research Ethics Committee (REF: PNM/12/13-102). Demographics are presented in Table 1 and the legend of Figure 8.

Table 1. Demographics from groups comparing cDNA allelic ratios from heterozygotes for risk variants versus respective gDNA ratios, as assayed per candidate gene.

Sample	Heterozygous for risk variant	*BORCS7*	*AS3MT*	*CNNM2*	*NT5C2*
Adult DLPFC	rs11191419	19M, 9F 71.6 (18–92)	6M, 2F 70.9 (54–90)	21M, 12F 72.8 (18–96)	20M, 11F 74.5 (18–96)
	chr10_104957618_I	8M, 1F 65.9 (18–88)	8M, 1F 67 (54–78)	11M, 5F 70.1 (18–96)	10M, 4F 71.3 (18–96)
Adult Hippocamp.	rs11191419	17M, 7F 76.5 (54–96)	5M, 2F 79.9 (67–92)	16M, 9F 77.8 (54–96)	13M, 10F 78.9 (54–96)
	chr10_104957618_I	7M, 4F 76.4 (54–96)	3M, 2F 78 (67–92)	9M, 5F 75.6 (54–96)	7M, 6F 77.5 (54–96)
Adult Caudate	rs11191419	12M, 10F 77.9 (42–96)	4M, 7F 78.8 (55–92)	18M, 12F 75.6 (18–96)	14M, 11F 74.9 (18–96)
	chr10_104957618_I	4M, 4F 81 (66–96)	3M, 3F 82 (69–92)	7M, 5F; 74.1 (18–96)	7M, 5F 74.1 (18–96)
Foetal brain	rs11191419	13M, 11F 106.3 (91–153)	7M, 10F 105 (91–153)	16M, 13F 113 (91–161)	10M, 10F 107.7 (91–153)
	chr10_104957618_I	4M, 4F 104.3 (91–133)	5M, 3F 107.4 (91–133)	3M, 3F 119.2 (93–154)	4M, 3F 111.4 (93–133)

Obs.: Values represent number of males (M), females (F), average age (range). Age is in years for adult samples, and in post-conception days for foetal samples. Adapted from Duarte et al. (2016).

2.3.2. Nucleic acids extraction (total RNA and genomic DNA)

In order to genotype samples for schizophrenia risk variants and the reporter SNPs used in the allele-specific expression assay, genomic DNA (gDNA) was isolated from all samples using phenol/chloroform liquid-liquid extraction. Briefly,

approximately 100 mg of frozen brain tissue was homogenised after overnight incubation at 50ºC in 1 mL of extraction buffer (NaCl 100 mM, TrisHCl 10 mM, EDTA 10 mM in 0.5% SDS in ultra-pure H_2O, pH 8) containing 6 units of proteinase K (ThermoFisher Scientific, Waltham, MA, USA). The solution was centrifuged (4 min at 4000 rpm, room temperature) and the supernatant mixed with 1 mL 1:1 mixture of chloroform:isoamyl alcohol (24:1) and phenol in a new tube. The sample was centrifuged and the supernatant diluted in 1 mL of chloroform:isoamyl alcohol (24:1). After harvesting, the aqueous layer containing nucleic acids was moved to a new tube and mixed with cold ethanol for centrifugation and precipitation. Ethanol was removed and the sample was let to air dry for 15 min. The pellet was resuspended in nuclease-free water and absorbances were measured using a NanoDrop ND1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA).

Total RNA was extracted by liquid-liquid extraction using Tri-Reagent (ThermoFisher Scientific). Approximately 100mg of frozen tissue was dissolved in Tri-Reagent by mechanical agitation (4 M/s, 40 seconds) in a FastPrep-24™ (MP Biomedicals, Santa Ana, CA, USA), in tubes containing Lysing Matrix D (MP Biomedicals). These tubes contain chemically inert ceramic beads which macerate the tissue upon mechanic agitation. Resulting homogenate was further mixed with 0.1 mL 1-bromo-3-phenolpropane per mL of Tri-Reagent used, and incubated at room temperature for 15 min. The aqueous phase was obtained upon centrifugation at 13.000 xg for 15 min at 4ºC, and mixed in a new tube with 0.5 mL isopropanol, molecular biology grade. This solution was centrifuged, the supernatant removed, and the pellet was washed three times with molecular grade 75% ethanol. Ethanol was removed, and the sample was let to air dry for 15 min. The pellet was resuspended in nuclease-free water and absorbances were measured using the ND1000 spectrophotometer. All reagents were purchased from Sigma (Sigma, St. Louis, MO, USA) unless stated otherwise.

2.3.3. DNAse treatment and cDNA synthesis

RNA samples were treated with TURBO DNA-free kit™ prior to cDNA synthesis, according to the protocol provided by the manufacturer. Briefly, 10 µL RNA (approximately 4 µg) was incubated with 2 units of Turbo DNAse in 1x Turbo DNAse buffer at 37ºC for 30 min. Reaction inactivation was performed by incubation with DNase Inactivation Reagent for 5 min at room temperature, instead of heat-inactivation, so that RNA was preserved from possible heat-induced degradation. The solution was centrifuged and the supernatant collected. This RNA did not yield a PCR product in the absence of a reverse transcription step.

Because the allele-specific expression assay is expected to control for RNA concentration (Bray et al., 2003b), reverse transcription does not need to occur with identical amounts of RNA. Therefore, cDNA was produced with approximately 1 µg of DNA-free RNA using SuperScript III® Reverse Transcriptase kit, according to the manufacturer's instruction. The reaction was performed in two steps that led to a final 20 µL volume reaction containing the equivalent cDNA, 5 µM random decamers, 500 µM dNTPs, 5 mM DTT, 40 units RNaseOUT™ Recombinant Ribonuclease Inhibitor and 200 units of reverse transcriptase enzyme in 1x First-strand buffer. An initial mixture of RNA, random decamers and dNTPs was incubated at 65ºC for 5 min (to remove secondary structures in the RNA), cooled on ice, and the remaining components were added. The reaction was incubated in a GS4 Thermocycler (G-Storm, Somerton, UK) for 5 min at 25ºC, 60 min at 50ºC, 30 min at 55ºC and 15 min at 70ºC, using a heated lid during the programme. The cDNA was diluted 1:7 and stored at -20ºC. Reagents were purchased from ThermoFisher Scientific.

2.3.4. Polymerase Chain Reaction (PCR) and PCR clean-up

Oligonucleotides were designed using Primer3 (Untergasser et al., 2012) and purchased from Integrated DNA Technologies (London, UK). Oligonucleotide sequences are shown on Table 2. PCR reactions were performed with HOT FIREPol® DNA Polymerase (Solis Biodyne, Tartu, Estonia) in 12 µL reactions containing 5-50 ng of cDNA or gDNA, 500 nM primers forward and reverse, 1× HOT FIREPol Buffer B1

containing 2.1 mM $MgCl_2$, 200 µM dNTPs and 0.5 unit of HOT FIREPol Taq Polymerase. The reaction was incubated in sealed 96-well plates in a GS4 Thermocycler (G-Storm) for 15 min at 95ºC (hot start), submitted to 35 cycles of 95ºC for 40 sec, 60ºC for 30 sec, 72ºC for 40 sec, and a final elongation step of 10 min at 72ºC, with a heated lid operating. Primers and nucleotides from PCR products were inactivated upon addition of 4 units of exonuclease I (ExoI) (New England Biolabs, Ipswich, MA, USA) and 1 unit of shrimp alkaline phosphatase (rSAP) (New England Biolabs). This reaction was incubated at 37ºC for 1 hour followed by an enzyme inactivation step of 85ºC for 15 min.

Table 2. Oligonucleotide sequences used in this study (Duarte et al., 2016).

Variant	MAF*	Usage	Forward and Reverse primers (5'-3')	Extension primer (5'-3')
rs4917985	0.37	Tag BORCS7	F: TGTTGTAGTTCAGGTCTTCATTGA R: TGCCTTTTCAGATCCTTCACA	TGCCTTTTCAGATCCTTCACATC
rs1046778	0.31	Tag AS3MT	F: CAAGAAGCAGGAAAGGCATC R: TTGGCAATCTTTTTGCATGA	GATCTTTTGCATAGCACCTT
rs2275271	0.39	Tag CNNM2	F: CGGGAACGAAAGCAAGATT R: GTTGCTAGGAAACGGTGCAT	TGTGGTGATATTTTAACCTTCATCTC
rs3740387	0.40	Tag NT5C2	F: CTCCCAACCTCTTCCCACTG R: GGACCTCGTTTGTTCCTGTG	CAGGAAATTACACACTGCCATGA
rs11191419	0.36	Risk SNP genotyping	F: TGAATCTAGACACTTGGAAGAGG R: GAAGTTGAGGCGGGAGGAT	CCACGCCCGGCTAATTTTTG
ch10_104957618_I	0.08	Risk indel sequencing	F: GAGTTAGGATCGGGTGAGGG R: GTCCGGTCATAGCTCACTGA	-

*MAF = minimum allele frequency.

2.3.5. Single nucleotide primer extension (SNaPshot®)

Genotyping of polymorphisms and the relative allelic expression were performed using SNaPshot® chemistry (ThermoFisher Scientific). The principle of single nucleotide extension applies to both cases, differing solely on the template used for the reaction (cDNA or gDNA). After PCR-amplifying a region of interest on the template, an extension primer that anneals just adjacent to the polymorphism of interest is extended by a single nucleotide using SNaPshot®. This reagent contains fluorescently labelled 2',3'-dideoxynucleotides (ddNTPs), which once incorporated cannot elongate due to the lack of a 3'-hydroxyl group, but provide a fluorescence signal for the incorporation.

The SNaPshot® assay consisted of 10 µL reactions that contained 10% SNaPshot® mix (v/v), 1 µM extension primer (final concentration) and 20% template

(v/v, clean PCR-product from cDNA or gDNA), diluted in ultra-pure water. The reaction was submitted to thermocycling in a GS4 Thermocycler (G-Storm) for 2 min at 95ºC and 30 cycles of 95ºC for 10 sec, 50ºC for 5 sec and 60ºC for 10 sec, with a heated lid during the programme. The SNaPshot product was diluted 1:4 in Hi-Di Formamide (ThermoFisher Scientific) for analysis in the ABI Genetic Analyzer 3130xl (ThermoFisher Scientific). Capillary electrophoresis coupled to a fluorescence detector enables the recording of peak heights representing the abundance of each fluorescently labelled allele, which can then be quantified using GeneMarker (SoftGenetics, State College, PA, USA). Examples of output from this allele-specific expression assay for a C/T polymorphism are given on Figure 2.

Samples were first genotyped for the schizophrenia risk SNP rs11191419 and for the reporter SNPs rs4917985, rs1046778, rs2275271 and rs3740387, which were used to tag *BORCS7, AS3MT, CNNM2* and *NT5C2,* respectively (Table 2). There was no significant deviation from Hardy-Weinberg equilibrium in the distribution of genotypes in any case. The expressed polymorphisms were then used to infer allelic expression in heterozygotes for these reporter SNPs. Four technical replicates for each cDNA and gDNA sample were assayed for the reporter SNPs. Samples showing poor reproducibility between technical replicates (standard deviation / mean > 0.25) were excluded from analysis. For each plate, the average allele ratio from all gDNA samples was used as a correction factor, since this can be assumed to reflect a perfect 1:1 ratio and can therefore control for inequalities in allelic representation inherent to the assay (Bray et al., 2003b).

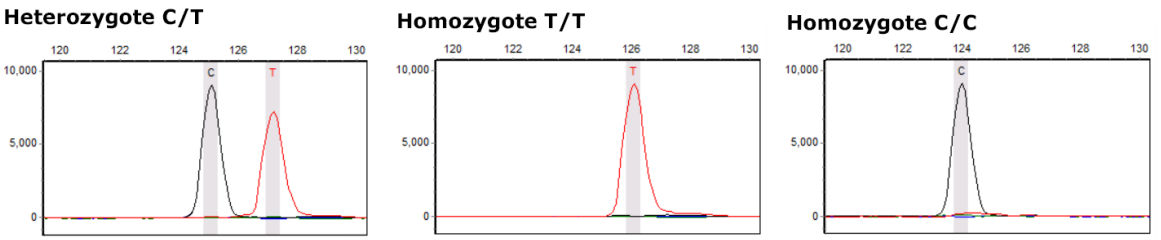


Figure 2. Example of genotyping output for a C/T polymorphism using SNaPshot. The reaction product was analysed in the ABI Genetic Analyzer 3130xl. Graphs were generated using GeneMarker (Softgenetics). Reactions can be performed using gDNA (for genotyping) or cDNA (for allelic expression quantification). Allelic expression can be calculated for heterozygotes for the reporter SNPs by dividing peak heights, as obtained from a SNaPshot reaction performed with cDNA. Y-axis: fluorescence intensity; X-axis: electrophoresis run time, in seconds.

2.3.6. Sanger sequencing (BigDye Terminator v3.1)

Samples were genotyped for the schizophrenia risk indel ch10_104957618_I (rs202213518) using BigDye Terminator v3.1 (ThermoFisher Scientific), as primers designed for SNaPshot reaction contained unacceptable GC content and did not yield specific products. BigDye contains a mixture of dNTPs and fluorescent ddNTPs which enables the selective incorporation of chain-terminating 2',3'-dideoxynucleotides by a DNA polymerase during amplification. By analysing the reaction product under capillary electrophoresis coupled to a fluorescence detector it is possible to sequence short fragments of DNA. Sequencing was performed in both directions, and a consensus sequence was obtained. Briefly, clean PCR product (15%, v/v) was incubated with 5% Big Dye Mix (v/v) and 500 nM primer forward or reverse in 1x BigDye buffer in a 10 µL reaction diluted with ultra-pure water. The reaction was thermocycled with a heated lid in the G-Storm GS4, with an initial temperature of 96ºC for 1 min, followed by 15 cycles of 96ºC for 8 sec, 50ºC for 8 sec, 60ºC for 90 sec, 5 cycles of 96ºC for 8 sec, 50ºC for 8 sec, 60ºC for 105 sec, 5 cycles of 96ºC for 8 sec, 50ºC for 8 sec and 60ºC for 120 sec, and a final 10ºC step for 10 min.

The sequencing product was precipitated using EDTA 7.4 mM (final concentration) and ethanol 70% (v/v, final concentration) by incubation at room temperature for 15 min, protected from light. Samples were centrifuged at 2,500 xg for 30 min at room temperature and wells were washed with 70% ethanol. After centrifugation, DNA was re-solubilised in 10 µL HiDi™ Formamide (ThermoFisher Scientific), which prevents formation of secondary structures during the capillary electrophoresis. Samples were analysed in the ABI 3130xl Genetic Analyzer (ThermoFisher Scientific) and the resulting electropherograms were processed on the ABI Sequence Scanner Software v1.0 (Figure 3) (ThermoFisher Scientific).

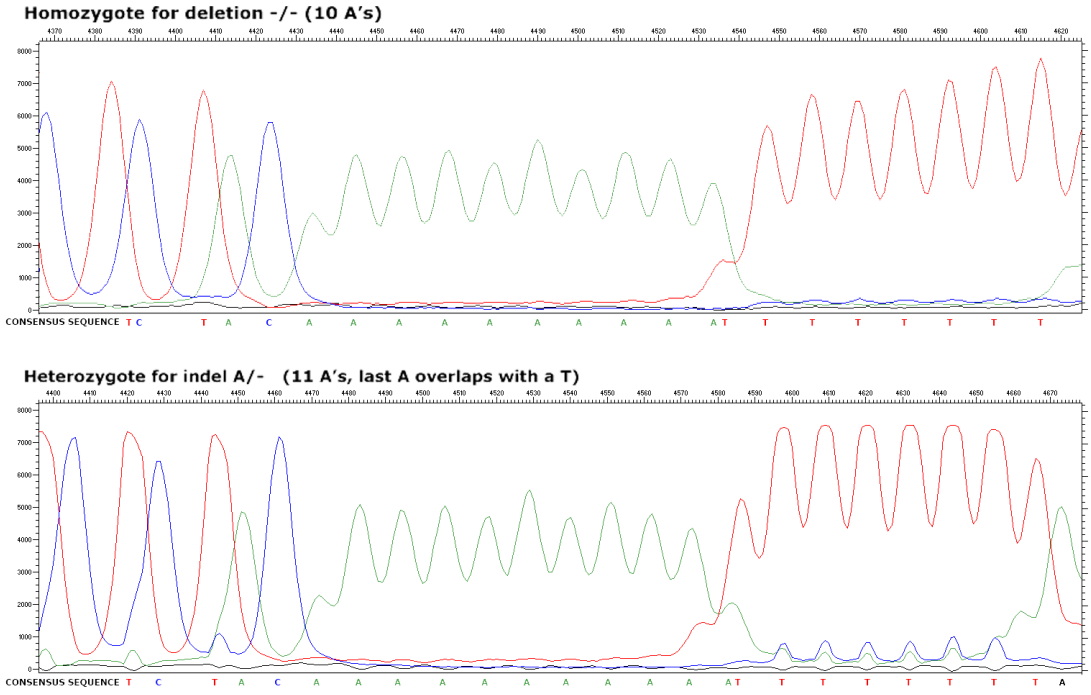


Figure 3. Sanger sequencing electropherograms for genotyping the risk indel. A homozygote (double deletion, top) and a heterozygote (insertion A/deletion, bottom) for ch10_104957618_I (rs202213518) are shown. X-axis indicates electrophoresis time, and Y-axis represents fluorescence intensity. No homozygote insertion/insertion (A/A) was found (MAF=0.08).

2.3.7. Association between schizophrenia risk alleles and allele-specific expression

Predicted haplotype frequencies were used to infer phase between reporter alleles and risk alleles of schizophrenia variants rs11191419 and ch10_104957618_I. Haplotypes were calculated by combining genotype data for risk variants and individual reporter SNPs from 116 adult and 95 foetal samples using Haploview 4.2 (Barrett et al., 2005). Haplotype counts estimated that risk alleles would nearly always be carried on the same chromosome as one of the reporter alleles for all genes due to strong linkage disequilibrium in the region. Because *cis*-regulatory effects act on the same chromosome as the regulatory variant, inferring phase (i.e. haplotype) allowed for the detection of the effect (up- or downregulation), as described elsewhere (Bray et al., 2005). This initial assessment was performed by comparing allelic ratios observed in gDNA (where there is no allelic imbalance) versus allelic imbalance in cDNA from subjects that were heterozygous for a risk variant. Allelic ratios were calculated in heterozygotes for the reporter SNP that were additionally heterozygous at the risk variants, by dividing the abundance of the reporter allele generally in phase with the schizophrenia risk alleles by the abundance of the reporter

allele generally in phase with the non-risk alleles. Comparisons were performed by two-tailed t-tests, and P-values < 0.05, Bonferroni-corrected for 32 tests, were considered significant, although nominal significance is also reported.

In samples that are heterozygous for the reporter SNP (due to the nature of this methodology) and additionally heterozygous for the risk variant (and thus double heterozygotes) it is necessary to determine phase between the risk allele and reporter alleles in order to determine whether the risk allele is associated with up- or downregulation of the gene. For these individuals, the frequency of the two possible diplotypes constructed from the two alleles of the reporter SNP and the two alleles of the risk variant was calculated, on the basis of predicted haplotype frequencies and the assumption of Hardy-Weinberg equilibrium, using the equations:

Frequency diplotype 1 = 2 X frequency haplotype A X frequency haplotype B

Frequency diplotype 2 = 2 X frequency haplotype C X frequency haplotype D

The probability that an individual who is heterozygous at both the reporter SNP and the risk variant is carrying diplotype 1 (comprising haplotypes A and B) is therefore calculated by dividing the predicted frequency of diplotype 1 by the combined frequency of both possible diplotypes. In order to test if fluctuations in the allelic expression imbalance could be accounted for by genotype at the risk variants, the cDNA allelic ratio in samples that were heterozygous for a risk variant (where *cis*-regulatory effects would cause differential allelic expression) was compared to that of cDNA from samples that were homozygous for that risk variant (where *cis*-regulatory effects would act equally on both chromosome strands) (Bray et al., 2003a; Bray et al., 2005; Hill and Bray, 2012). Given the low number of homozygotes observed for risk variant rs11191419, due to high r^2 between this and the reporter SNPs, these analyses were restricted to comparisons between homozygotes and heterozygotes for the risk indel. Comparisons were performed by two-tailed t-tests, and P-values < 0.05, Bonferroni-corrected for 16 tests, were considered significant, although nominal significance is also reported. All tests, performed in SPSS 22.0, were tested under an unequal variance model if a significant difference in variance was found between groups using the Levene's test (P < 0.05).

2.4. Results

Cis-effects on the expression of *BORCS7*, as measured by allelic expression abundance of the reporter SNP rs4917985 (Figure 4), were associated with genotype at risk SNP rs11191419. The reporter C-allele was found on the same chromosome as (in phase with) the T- risk allele of rs11191419 on >99% of occasions in samples that were heterozygous for the reporter and for the risk variant. The expression of the C-allele of this reporter SNP was additionally more abundant than its allelic counterpart in all brain regions (DLPFC: 12%, hippocampus: 8%, caudate: 11%, foetal brain: 5%). The observed allelic imbalance in the DLPFC was significantly different from the 1:1 ratio observed in gDNA (P=0.001, Bonferroni corrected), but only nominally significant in other brain areas (P<0.05, uncorrected).

The risk allele of the indel ch10_104957618_I (deletion) was predicted to be on the same chromosome as the reporter C-allele of *BORCS7* on >99% of occasions, but heterozygosity at this variant was not associated with a significant effect on the allelic expression of *BORCS7* in the assayed tissues.

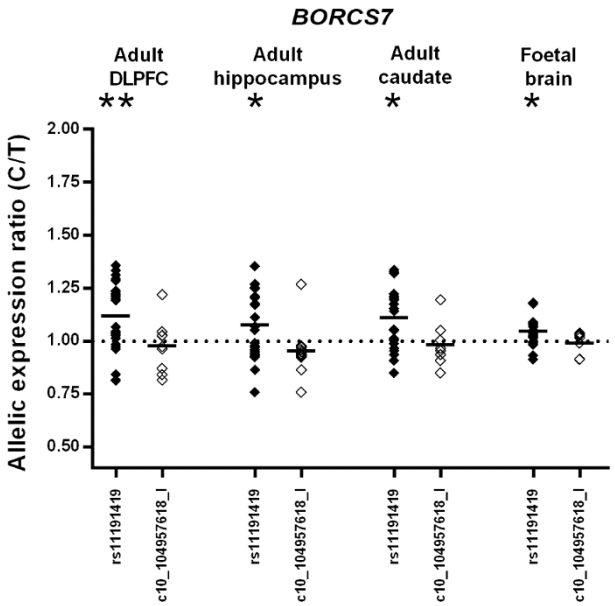


Figure 4. Allelic expression of *BORCS7* in heterozygotes for the schizophrenia risk variants. The C/T ratio (Y-axis) represents the abundance of the reporter allele generally in phase with the schizophrenia risk alleles by the abundance of the expressed allele generally in phase with the non-risk alleles. The average cDNA ratio for each group is represented by short bars. The dotted line represents the 1:1 allele ratio observed in gDNA, where no allelic imbalance occurs. The allelic imbalance observed in cDNA was compared to that of gDNA using two-tailed t-tests (*P< 0.05, uncorrected; **P<0.05, Bonferroni corrected for 32 tests). Extracted from Duarte et al. (2016).

Cis-regulatory effects on the expression of *AS3MT*, as measured by allelic expression abundance of the reporter SNP rs1046778 (Figure 5), was found significantly associated with SNP rs11191419 in the foetal brain. Nominally significant effects of both risk variants were also found in the hippocampus and caudate. The reporter T-allele was found in phase with the T- risk allele of rs11191419 on >94% of occasions, and was more abundant than its counterpart C-allele in all brain regions (DLPFC: 14%, hippocampus: 36%, caudate: 23%, foetal brain: 40%). The allelic expression imbalance in the foetal brain was significantly different from the 1:1 ratio observed in gDNA (P=0.00036, Bonferroni corrected), but only nominally significant in the hippocampus and caudate (P<0.05, uncorrected).

The risk allele of the indel ch10_104957618_I (deletion) was predicted to be on the same chromosome as the reporter T-allele on >99% of occasions. Heterozygosity for the risk allele was associated with moderately increased expression of the linked reporter allele in all tissues (DLPFC: 7%, hippocampus: 24%, caudate: 14%, foetal brain 14%). The allelic expression imbalance in the hippocampus, caudate and foetal brain differed from gDNA only with nominal significance (P<0.05, uncorrected).

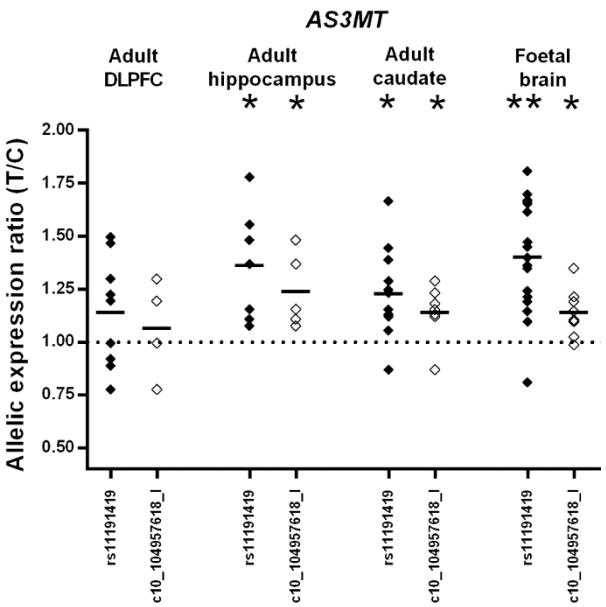


Figure 5. Allelic expression of *AS3MT* in heterozygotes for the schizophrenia risk variants. The T/C ratio (Y-axis) represents the abundance of the reporter allele generally in phase with the schizophrenia risk alleles by the abundance of the expressed allele generally in phase with the non-risk alleles. The average cDNA ratio for each group is represented by short bars. The dotted line represents the 1:1 allele ratio observed in gDNA, where no allelic imbalance occurs. The allelic imbalance observed in cDNA was compared to that of gDNA using two-tailed t-tests (*P< 0.05, uncorrected; **P<0.05, Bonferroni corrected for 32 tests). Extracted from Duarte et al. (2016).

Cis-regulatory effects on the expression of *CNNM2*, as measured by allelic expression abundance of the reporter SNP rs2275271 (Figure 6), did not show a strong association with any of the studied risk variants. The reporter A-allele was found in phase with the T- risk allele of rs11191419 on >98% of occasions, but unlike *BORCS7* and *AS3MT*, cDNA allele ratios in heterozygotes for rs11191419 were close to the 1:1 ratio. A 5% decrease in the mean expression of the reporter A-allele was observed in phase with the risk allele, differing only nominally in the foetal brain (P<0.05, uncorrected).

The risk allele of the indel ch10_104957618_I (deletion) was predicted to be in phase with the reporter T-allele on >99% of occasions, and was associated with a small (5%) decrease in its expression in the adult DLPFC, which was nominally significant (P<0.05, uncorrected). Allelic expression imbalance of *CNNM2* in association with these schizophrenia risk variants did not survive Bonferroni correction in any region.

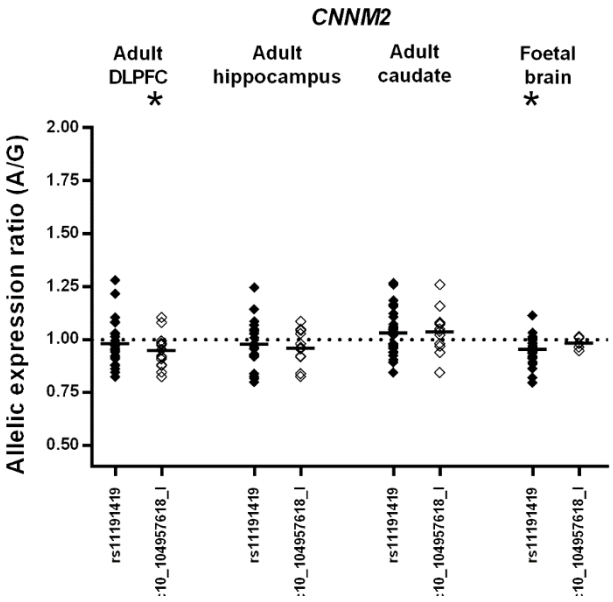


Figure 6. Allelic expression of *CNNM2* in heterozygotes for the schizophrenia risk variants. The A/G ratio (Y-axis) represents the abundance of the reporter allele generally in phase with the schizophrenia risk alleles by the abundance of the expressed allele generally in phase with the non-risk alleles. The average cDNA ratio for each group is represented by short bars. The dotted line represents the 1:1 allele ratio observed in gDNA, where no allelic imbalance occurs. The allelic imbalance observed in cDNA was compared to that of gDNA using two-tailed t-tests (*P< 0.05, uncorrected; **P<0.05, Bonferroni corrected for 32 tests). Extracted from Duarte et al. (2016).

Cis-regulatory effects on the expression of *NT5C2*, as measured by allelic expression abundance of the reporter SNP rs3740387 (Figure 7), was found to be significantly associated with both risk variants in three adult brain regions, and nominally significant in the foetal brain. The reporter C-allele was found in phase with the T- risk allele of rs11191419 on >98% of occasions. The reporter C-allele was less abundant than its counterpart in all adult brain regions (mean DLPFC: 11%, hippocampus: 7%, caudate: 7%), with cDNA allele ratios in DLPFC and caudate differing significantly from the ratio observed in genomic DNA (respectively, $P=2.01\times10^{-5}$ and P=0.032, Bonferroni corrected), but only nominally significant in the hippocampus (P<0.05, uncorrected).

The risk allele of the indel ch10_104957618_I (deletion) was predicted to be on the same chromosome as the reporter C-allele on >99% of occasions, which was found to be less represented than its allelic counterpart in all tested brain regions (mean DLPFC: 15%, hippocampus: 12%, caudate: 13%, foetal brain: 7%). Allelic expression ratios for *NT5C2* in the DLPFC, hippocampus and caudate differed significantly from the ratios observed in genomic DNA (respectively, *P*=0.003, $P=8.46\times10^{-5}$ and *P*=0.0025, Bonferroni-corrected), but only nominally in the foetal brain (P<0.05, uncorrected).

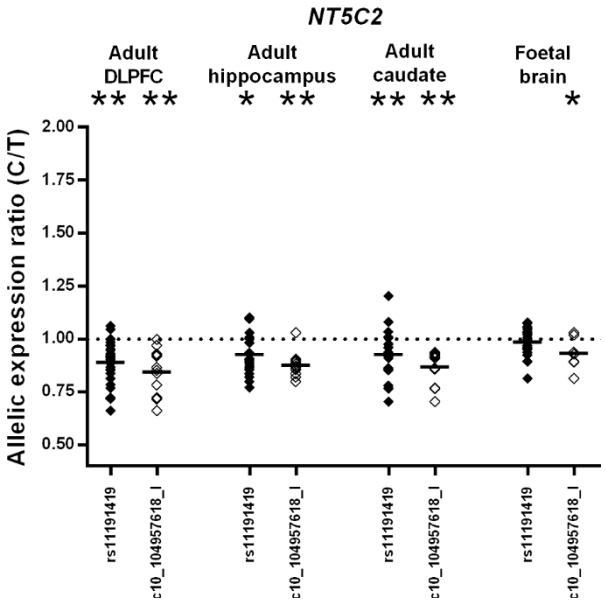


Figure 7. Allelic expression of *NT5C2* in heterozygotes for the schizophrenia risk variants. The C/T ratio (Y-axis) represents the abundance of the reporter allele generally in phase with the schizophrenia risk alleles by the abundance of the expressed allele generally in phase with the non-risk alleles. The average cDNA ratio for each group is represented by short bars. The dotted line represents the 1:1 allele ratio observed in gDNA, where no allelic imbalance occurs. The allelic imbalance observed in cDNA was compared to that of gDNA using two-tailed t-tests (*P< 0.05, uncorrected; **P<0.05, Bonferroni corrected for 32 tests). From Duarte et al. (2016).

In order to investigate whether genotype at the risk variant ch10_104957618_I could account for the observed *cis*-regulatory mechanisms, allelic expression imbalance for each gene was compared between heterozygotes and homozygotes for this risk variant. This investigation was performed only for the risk indel ch10_104957618_I, as the small number of homozygotes for rs11191419 precluded this group from statistical analysis. *Cis*-regulatory mechanisms attributed to genotype at ch10_104957618_I (Table 3) included effects on *BORCS7* in the adult hippocampus (P=0.016, corrected) and foetal brain (P=0.032, corrected), *AS3MT* in the foetal brain (P=0.032, corrected), and *NT5C2* in the adult DLPFC (P=0.048, corrected). Additionally, the risk indel was found to be associated with a small reduction in the expression of *BORCS7* and *AS3MT*, which was being increased by the risk allele of rs11191419, consistent with the observations in Figures 4 and 5. The risk alleles of rs11191419 and ch10_104957618_I were also found to largely account for the reduced expression of the reporter C-allele of *NT5C2* in the DLPFC (mean 16% reduction), with homozygotes for these risk variants displaying allelic expression ratios close to the observed in gDNA (Figure 8). The allelic expression ratios of *NT5C2* in heterozygotes for the risk variants statistically differed from homozygotes for the risk variants (P=0.007) and from gDNA (P<0.001). In other words, both risk variants exerted opposing effects on *AS3MT* and *BORCS7,* whereas they exerted concordant effects on *NT5C2*.

Table 3. Average allelic expression of *BORCS7*, *AS3MT*, *CNNM2* and *NT5C2,* according to genotype at the schizophrenia risk indel.

Gene (expressed SNP)[#]	Genotype ch10_104957618_I	Adult DLPFC	Adult hippocampus	Adult caudate	Foetal whole brain
BORCS7 (rs4917985) C/T	Heterozygous	0.99	0.95	0.98	0.99
	Homozygous	1.17	1.12	1.15	1.06
	P het versus hom	**0.006**	**0.001***	**0.004**	**0.002***
AS3MT (rs1046778) T/C	Heterozygous	1.07	1.24	1.14	1.14
	Homozygous	1.25	1.52	1.31	1.42
	P het versus hom	0.235	0.187	**0.034**	**0.002***
CNNM2 (rs2275271) A/G	Heterozygous	0.95	0.96	1.04	0.98
	Homozygous	0.99	0.98	1.02	0.95
	P het versus hom	0.095	0.629	0.527	0.342
NT5C2 (rs3740387) C/T	Heterozygous	0.85	0.88	0.87	0.93
	Homozygous	0.94	0.95	0.93	0.99
	P het versus hom	**0.003***	**0.014**	0.072	**0.016**

Allele ratios at each expressed SNP were calculated by dividing measures of the allele generally in phase with the schizophrenia risk alleles by measures of the allele generally in phase with the non-risk alleles, as indicated.
***** P-values that survive Bonferroni correction for 16 tests. Uncorrected P-values < 0.05 are indicated in bold.
Extracted from Duarte et al. (2016).

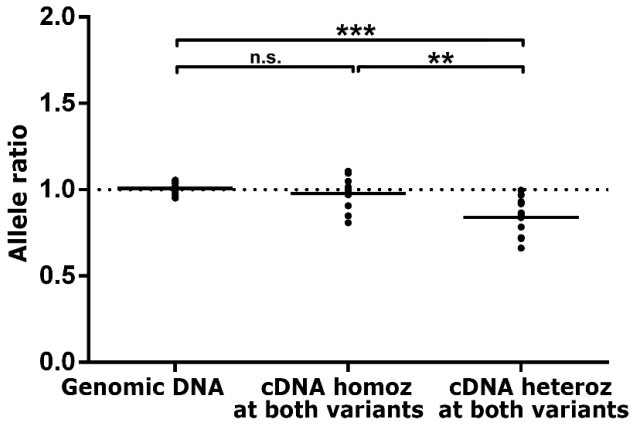


Figure 8. Effect of risk genotype in the allelic expression of *NT5C2* in the DLPFC. Allelic expression ratios are shown in the DLPFC of adult subjects who are homozygous for both risk variants rs11191419 and ch10_104957618_I (6 males, 3 females, average age of 64 years), and heterozygous for both variants (9 males, 4 females, average age of 70 years). Sharp horizontal lines indicate mean allelic ratio. The dotted line represents the mean genomic DNA (1:1) ratio where no cis-regulatory effects operate. Allele expression ratio from subjects who are homozygotes for both risk variants do not differ from allelic ratio in genomic DNA, but differ significantly from the allelic expression ratio observed in heterozygotes for both variants. **P<0.01, ***P<0.001. Extracted from Duarte et al. (2016).

2.5. Discussion

More than 90% of the schizophrenia susceptibility variants implicated by GWAS are non-coding, and are therefore likely to impact on gene expression (Schizophrenia Working Group of the PGC, 2014). These variants, however, frequently correlate in genotype with neighbouring variants due to LD, causing association signals to span multiple genes. For instance, the latest schizophrenia GWAS found 108 independent genomic loci to be associated with susceptibility. From these, because of linkage disequilibrium, 107 signals comprise of over 350 genes (Schizophrenia Working Group of the PGC, 2014), and one signal, at the MHC region on chromosome 6, spans over hundreds of genes by itself (Sekar et al., 2016). It is necessary, therefore, to couple measures of gene expression to these genetic findings, in order to resolve the association signals and expose potential risk mechanisms implicated in disease susceptibility (Bray and Hill, 2016).

Chromosome 10q24 is the third most significant risk locus from the largest genome-wide association study of schizophrenia to date (Schizophrenia Working Group of the PGC, 2014), for which association with this disorder has been extensively replicated (Aberg et al., 2013; Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011). Schizophrenia risk SNPs at the locus (and variants in LD with them) have been associated with an index of social cognition and with brain morphometric differences implicated in schizophrenia (Ohi et al., 2013; Ohi, 2015; Rose et al., 2014), but the mechanisms driving association with the disease remain unknown. In this context, this chapter provides functional evidence to support *BORCS7, AS3MT* and *NT5C2,* as putative risk genes for schizophrenia. The two most supported risk variants can be found flanking the LD region (Figure 1), with risk variant rs11191419 located 2kb upstream of the transcriptional start site of *BORCS7* (plus strand), and the risk indel ch10_104957618_I located 4.5kb upstream of the transcriptional start site of *NT5C2* (minus strand). More specifically, the indel is located in a H3-K27ac-marked region according to ENCODE (Chip-seq data), denoting a regulatory binding site for multiple

transcription factors. Findings from this chapter, in summary, suggest that both risk variants have converging *cis*-regulatory effects on *NT5C2* by reducing its expression in several brain regions, but elicit opposing effects on the expression of *BORCS7* and *AS3MT.* The risk allele of rs11191419 was found to be associated with increased expression of *BORCS7* and *AS3MT*, while the risk allele of ch10_104957618_I was found to reduce their expression, partially counteracting the effect of rs11191419.

The observed pattern of *cis*-regulatory mechanisms operating on chromosome 10q24 is consistent with the analysis of eQTL (expression quantitative trait loci) and *cis*-regulatory elements (CREs) reported by Roussos et al. (2014). The authors found that rs7085104 (which is in strong LD with rs11191419: r^2=0.79) influenced expression of *BORCS7, AS3MT, WBP1L* and *NT5C2* in a combined analysis of several brain eQTL databases. Consistent with data generated in this chapter, they also showed that *BORCS7* (synonym: *C10ORF32*) and *AS3MT* are part of a single CRE unit, whereas *NT5C2* was involved with at least 14 CREs.

Overexpression of *AS3MT* in blood was found to be associated with genotype at rs7096169 (which is also in strong LD with rs11191419: r^2=0.85) (Schizophrenia Working Group of the PGC, 2014). Moreover, differential methylation at several CpG islands spanning *AS3MT* was found to be associated with rs11191419 in human foetal brain (Hannon et al., 2016), consistent with regulatory mechanisms operating on this gene early in life. A recent study found an increased expression of *BORCS7* and *AS3MT* in association with genotype at rs7085104 (which is in strong LD with rs11191419: r^2=0.79) in several brain regions (Li et al., 2016), supporting the findings presented in this chapter. Furthermore, this risk SNP was observed to tag two other genetic variants at the locus, which are distinctively regulating expression of these genes: rs11441374, which drives expression of BORCS7; and a variable number tandem repeat (VNTR) in exon 1 of *AS3MT,* which regulates expression of this gene and of a new truncated transcript (*AS3MT^{d2d3}*), that the authors propose as the true schizophrenia risk transcript of *AS3MT*. Interestingly, the novel risk

transcript produces a protein lacking the methyltransferase domain, suggesting that the encoded protein is not involved in arsenic metabolism.

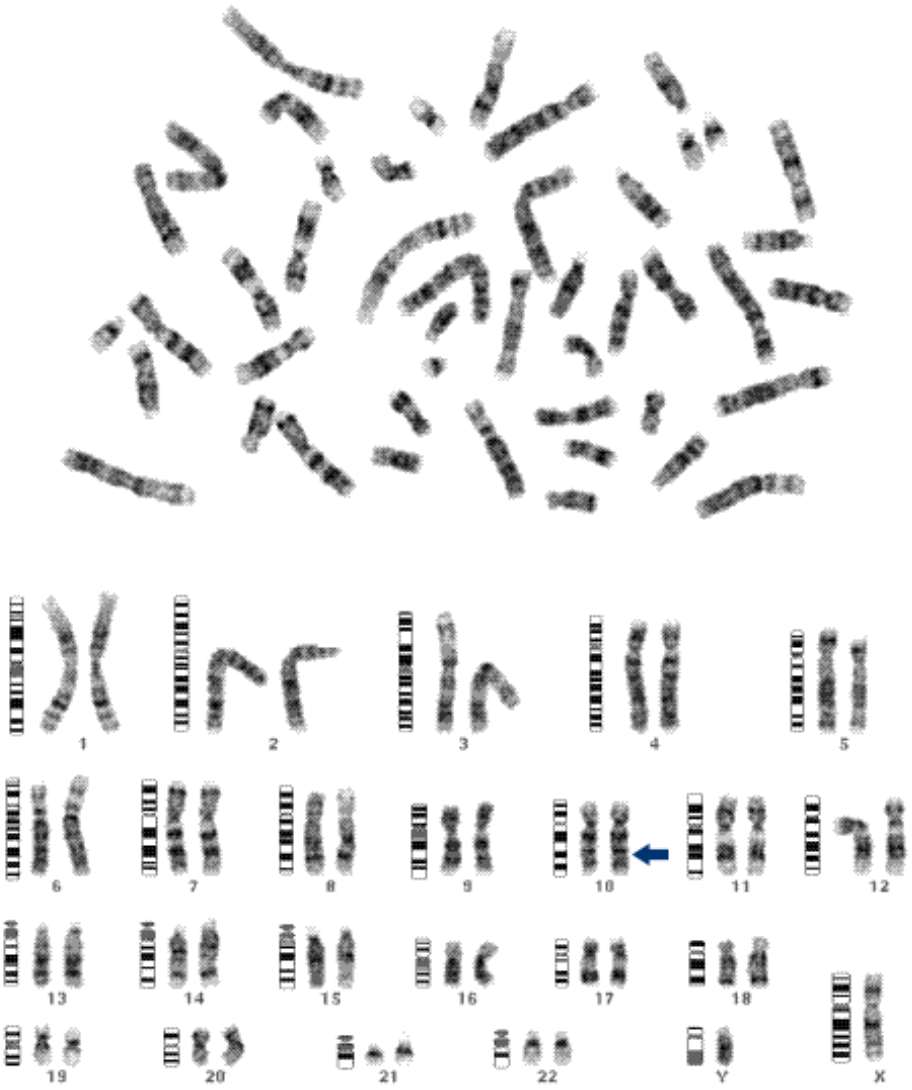
The *cis*-regulatory mechanisms operating on *NT5C2*, however, were associated with both risk alleles causing decreased expression of this gene in all brain areas assessed. This suggests that risk to schizophrenia on chromosome 10q24 is likely conferred via decreased expression of *NT5C2* in multiple brain tissues or developmental time points. Strikingly, genotype at the risk variants accounted for the majority of *cis*-regulatory effects observed for this gene in the adult DLPFC (Figure 8), substantiating the choice for focusing on this gene in the next chapters. Li et al. (2016) did not observe an effect of rs11191419 on *NT5C2* expression, likely because their eQTL-based approach is expected to be less sensitive for detecting the small *cis*-regulatory effects observed for this gene. In this chapter, regardless, it is noteworthy that there is a striking consistency of the effects of the risk alleles on *BORCS7, AS3MT,* and *NT5C2,* across brain tissues, which is particularly remarkable given that these analyses were subject to tissue availability and reporter allele frequencies. Moreover, a recent study suggested that the schizophrenia protective allele of rs11191548 (which is in strong LD with risk indel ch10_104957618_I: r^2=0.82) disrupts binding of microRNA-206 to this variant, located on the 3' UTR (untranslated region) of *NT5C2,* leading to increased *NT5C2* transcript availability (Hauberg et al., 2016), and therefore corroborates our findings.

Public databases indicate that BORCS7, *AS3MT* and *NT5C2* are differentially expressed throughout different maturation stages of the human brain (Birnbaum et al., 2015; Kang et al., 2011). The *cis*-regulatory mechanisms operating on these genes during early development supports the neurodevelopmental hypothesis of schizophrenia (Murray and Lewis, 1987; Weinberger, 1987). Interestingly, these effects persist in several brain regions later in adulthood, which suggest that these might be important ongoing events. The functions of the assayed genes remain to be elucidated in the context of schizophrenia. *BORCS7* was described recently as encoding the BLOC-1 related complex subunit 7 (diaskedin), which is implicated in

lysosomal function and cell migration (Pu et al., 2015). *AS3MT* encodes an arsenic methyltransferase, enzyme involved in metabolism of arsenic compounds (Sumi and Himeno, 2012). *NT5C2,* in turn, encodes the cytosolic 5'-nucleotidase II (or cN-II), which is involved in the metabolism of intracellular purine nucleotides such as IMP, AMP and GMP (Itoh, 2013), which therefore support the purinergic hypothesis of schizophrenia. Curiously, this theory covers several aspects of the disorder, including neurochemical findings, their association with positive and negative symptoms, and the developmental component of schizophrenia (Lara and Souza, 2000).

This study does not exclude possible *cis*-regulatory effects elicited on genes other than those tested due to long range *cis*-regulatory effects, or developmental stage- or brain region-specific events. More brain regions could be analysed in order to provide a more comprehensive assessment. More sophisticated statistical analyses and study designs, for example summary data-based Mendelian randomization (SMR) (Zhu et al., 2016) and transcriptome-wide association studies (TWAS) (Gusev et al., 2016; Pirinen et al., 2015; Smith et al., 2013), could be adopted to increase confidence that risk is operating through regulatory effects on particular genes. It is also possible that susceptibility to schizophrenia is conferred by specific or novel transcript variants of the implicated genes, as the allelic expression signals captured in this study are product of all transcripts containing the reporter SNPs. In this scenario, RNA-seq is the tool of choice for detecting all transcripts expressed in a given tissue, which can be further explored using other molecular tools such as rapid amplification of cDNA ends (RACE) or RT-qPCR (as performed in Tao et al., 2014).

This chapter comprises the first ever published study to analyse *cis*-regulatory mechanisms associated with schizophrenia risk variants on chromosome 10q24 (Duarte et al., 2016). The risk polymorphisms were associated with altered *cis*-regulation of *BORCS7, AS3MT* and *NT5C2*, therefore suggesting these as putative risk genes for schizophrenia. The following chapters will investigate these mechanisms in more detail, with particular emphasis on *NT5C2*.



The 22 autosomal chromosome pairs and the sex chromosomes Y and X. Approximate location of cytoband 10q24 is indicated by an arrow. Image adapted from Schreck and Disteche (2001).

Chapter 3

Chromosome 10q24 transcript characterisation in the human brain and neural cell lines

3.1. Summary

In the previous chapter, schizophrenia risk variants on chromosome 10q24 were found to be associated with *cis*-regulatory effects on *BORCS7, AS3MT* and *NT5C2,* in the dorsolateral prefrontal cortex (DLPFC), hippocampus, caudate and, to some extent, in the foetal brain. It remains unclear which transcripts of these genes are produced in these brain tissues, contributing towards the observed allelic expression imbalances. In this context, a pilot RNA sequencing (RNA-seq) study was designed to determine all transcripts produced by these genes in the adult DLPFC and the second trimester foetal brain. Manual analysis of junction reads provided evidence for novel splicing variants of *NT5C2* and *AS3MT,* including the recently proposed schizophrenia susceptibility transcript $AS3MT^{d2d3}$, in these brain samples. Given the underpowered nature of this pilot experiment, these non-annotated transcripts were not significantly detected, according to the Tuxedo pipeline. However, they were experimentally validated in multiple samples of the foetal brain, adult DLPFC and other adult brain areas using RTPCR and transcript-specific primers. No evidence for novel *BORCS7* transcripts was observed, although expression of its RefSeq transcript NM_001136200 was found to be higher than the alternative NM_144591 in the RNA-seq data. Novel less expressed *NT5C2* transcripts were found, which are denoted here as $NT5C2^{d3}$, $NT5C2^{e3.1}$ and $NT5C2^{e3.1\ d3}$. Expression of the *NT5C2* transcript NM_001134373 was also found to be higher than expression of the alternative variant NM_012229. These data provide evidence to support the existence of novel transcripts of *AS3MT* and *NT5C2,* although RNA-seq data suggest that their full length counterparts are more highly expressed in the assayed brain regions.

Note: Unpublished expression microarray data from neuronal differentiation in cortical and hippocampal neural progenitor cells were provided by Dr Timothy Powell, Dr Sandrine Thuret, Dr Gerome Breen, Dr Deepak Srivastava, Dr Nick Bray and Dr Greg Anderson.

3.2. Introduction

One of the most robust signals to emerge from large scale schizophrenia genome-wide association studies (GWAS) is on chromosome 10q24 (Aberg et al., 2013; Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Schizophrenia Working Group of the PGC, 2014). In Chapter 2, the top risk variants at this locus were associated with altered *cis*-regulatory effects on the focal candidates *BORCS7, AS3MT* and *NT5C2* (Figure 1). The methodology employed to assess these *cis*-regulatory events, however, precluded from identifying specific transcripts involved in the risk mechanism, as the allelic expression measurement was averaged for all transcripts containing the reporter SNPs. *Cis*-regulatory mechanisms may confer risk to schizophrenia via specific or non-annotated gene transcripts, as has been suggested for *ZNF804A* (*ZNF804A*^{E3E4}) (Tao et al., 2014). Therefore, the objective of this chapter is to perform a pilot RNA-seq study to identify chromosome 10q24 transcripts produced in brain regions where cis-regulatory effects associated with schizophrenia were identified in Chapter 2.

RNA-seq actually refers to high-throughput cDNA sequencing, and it is a powerful tool for transcriptome profiling. Several library preparation kits and various next-generation sequencing (NGS) platforms, using different chemistry sequencing methods, are commercially available. These include the Illumina HiSeq and MiSeq, Qiagen, SOLiD, Ion Torrent, Roche 454 (Goodwin et al., 2016), and also the more recent 'third generation sequencing' methods, capable of sequencing large stretches of DNA, such as the Oxford Nanopore Technologies, the Pacific Biosciences Single-Molecule Real-Time and the Illumina Tru-Seq Synthetic Long-Read Technology (Lee et al., 2016). Furthermore, analysis of data produced by these platforms is currently subject of intensive research in Bioinformatics, and there is a plethora of tools available to suit user's needs (Milicchio et al., 2016).

In this chapter, RNA-seq was performed using the Illumina HiSeq 2500 and the data was analysed using the Tuxedo suite (Trapnell et al., 2012). Briefly, Illumina paired-end reads were mapped to the reference genome using Tophat2/Bowtie2 (Kim et al., 2013; Langmead and Salzberg, 2012). The reads were assembled into transfrags (i.e. short transcribed fragments in the sequencing data) using Cufflinks (Trapnell et al., 2010), and these were assembled into transcripts using Cuffmerge (Trapnell et al., 2010). These results were used to direct the design of transcript-specific primers for reverse transcription PCR (RT-PCR). This quick and robust method was used to validate RNA-seq findings and to investigate expression of relevant transcripts in multiple brain regions. This pilot study analysed an RNA sample from the *post-mortem* adult dorsolateral prefrontal cortex (DLPFC) and the second trimester foetal brain, where *cis*-regulatory effects on chromosome 10q24 were observed in association with genotype at schizophrenia risk variants.

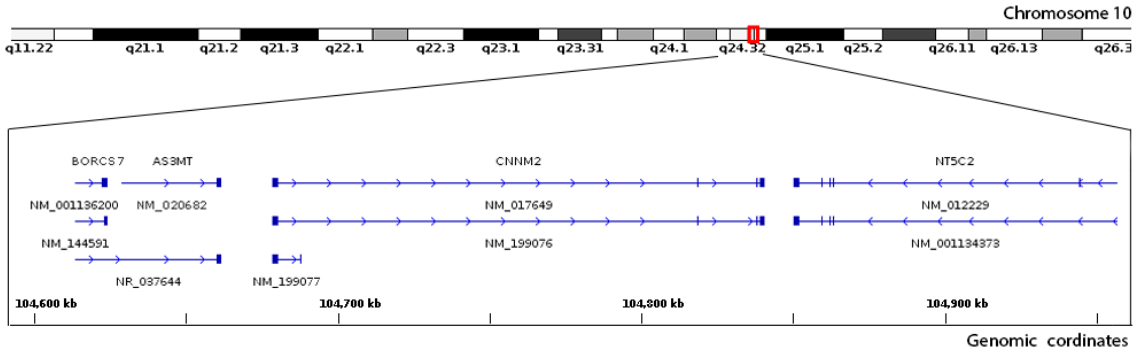


Figure 1. RefSeq transcripts on chromosome 10q24. Generated by IGV Browser (Broad Institute) using the genome built hg19. Chromosome 10 locus q24.32-33 is a region implicated in risk for schizophrenia, where linkage disequilibrium hinders the identification of the true schizophrenia susceptibility genes (Chapter 2, Figure 1). Although *BORCS7, AS3MT* and *NT5C2* were implicated as putative schizophrenia risk genes in Chapter 2, it remains unclear which transcript(s) of these genes is(are) conferring risk for mental illness.

3.3. Methods

3.3.1. Brain samples

Human *post-mortem* dorsolateral prefrontal cortex tissue from a 54 year-old male subject (sample A130/09), free from psychiatric or neurological diagnosis at the time of death, was obtained from the MRC London Neurodegenerative Diseases Brain Bank at the Institute of Psychiatry, Psychology and Neuroscience, King's College London (UK). Whole foetal brain sample from a 94 post-conception day female foetus (sample 15533) resulting from an elective abortion was obtained from the MRC Wellcome Trust Human Developmental Biology Resource (HDBR) at the Institute of Child Health, University College London (UK). The mother's consent for research, including genetic analysis, was obtained by the HDBR. Validation of RNA-seq findings by RT-PCR was carried out on additional second trimester foetuses and adult dorsolateral prefrontal cortex, hippocampus and caudate (Table 1). Ethical approval for this study was provided by The Joint South London and Maudsley and The Institute of Psychiatry NHS Research Ethics Committee (reference PNM/12/13-102).

Table 1. Demographics from samples used for RT-PCR.

Tissue	Mean age (range)*	Males (M) / Females (F)
DLPFC	78.3 (54-96)	1M / 3F
Hippocampus	66 (54-89)	1M / 3F
Caudate	64.5 (41-92)	2M / 2F
Second trimester whole foetal brain	96 (93-104)	2M / 2F

* The age of adults is given in years, and of foetuses in post-conception days.

3.3.2. RNA extraction

Total RNA was extracted as in Chapter 2. Briefly, approximately 100 mg of frozen tissue was dissolved in 1 mL Tri-Reagent (ThermoFisher Scientific, Waltham, MA, USA) with mechanical agitation at 4 M/s for 40 sec in a FastPrep-24™ (MP Biomedicals, Santa Ana, CA, USA) in tubes containing Lysing Matrix D (MP Biomedicals). The solution was mixed with 0.1 mL 1-bromo-3-phenolpropane and incubated at room temperature for 15 min. The sample was centrifuged at 13.000 xg for 15 min, at 4ºC, and the aqueous phase was transferred to new tube containing 0.5 mL isopropanol. After centrifugation, the supernatant was removed, and the

pellet was washed three times with molecular grade 75% ethanol. Residual ethanol was removed, and the pellet was air dried for 15 min. The pellet was resuspended in nuclease-free water and absorbance ratios were calculated using an ND-1000 spectrophotometer. All reagents, including the nuclease-free water, were purchased from Sigma, unless stated otherwise.

3.3.3. DNAse treatment and RNA integrity

Samples were treated with the TURBO DNA-free kit™, as described in Chapter 2. Briefly, 10 µL RNA (approximately 4 µg) was incubated with 2 units of Turbo DNAse in 1x Turbo DNAse buffer at 37ºC for 30 min. Reaction inactivation was performed by incubation with DNase Inactivation Reagent for 5 min. Samples were harvested and the supernatant was collected. This DNA-free RNA did not yield a PCR product in an agarose gel. For samples submitted to RNA-seq, RNA integrity was assessed using the Agilent RNA 6000 Nano Kit on an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). The RNA integrity number (RIN) was >7 for both samples and their respective values were used to dictate the shearing step in the library preparation protocol.

3.3.4. Library preparation and RNA sequencing

The RNA-seq libraries were constructed using the TruSeq RNA Library Prep Kit v3 (Illumina, San Diego, CA, USA), according to the manufacturer's instructions. This protocol generates a library of cDNA molecules with Illumina adapters ligated to each end, suitable for sequencing in Illumina platforms. First step of this process involved the depletion of ribosomal RNA (rRNA) from total RNA by using biotinylated target-specific oligos, and Ribo-Zero Gold rRNA Removal Beads that target mitochondrial and cytoplasmic rRNA. Depleted RNA was fragmented in a buffer containing divalent cations and incubated at high temperature for the period of time suggested by the manufacturer, according to sample's RIN (degraded samples need less shearing time than more conserved samples). Fragmented RNA was submitted to first strand cDNA synthesis using reverse transcriptase, and, subsequently, second

strand synthesis using DNA Polymerase I and RNAse H for, respectively, cDNA synthesis and removal of the RNA strand. The polymerase adds an extra "A" nucleotide, which was later used for the ligation of paired-end adapters. The reaction product was purified and PCR-amplified using primers specific to the adapters. Quality of the RNA-seq libraries was assessed in an Agilent High Sensitivity DNA chip using the Agilent 2100 Bioanalyzer (Agilent), to determine obtained average fragment length (Figure 2). Sequencing was performed in single lanes of the HiSeq 2500 Ultra-High-Throughput Sequencing System (Illumina), at the Exeter Sequencing Service at the University of Exeter. This sequencing technology relies on the high-throughput synthesis of DNA molecules that are complementary to templates in the library. Each added base releases a fluorescent signal that is detected by the sequencer. Raw FASTQ files were obtained and analysed as described in the next section.

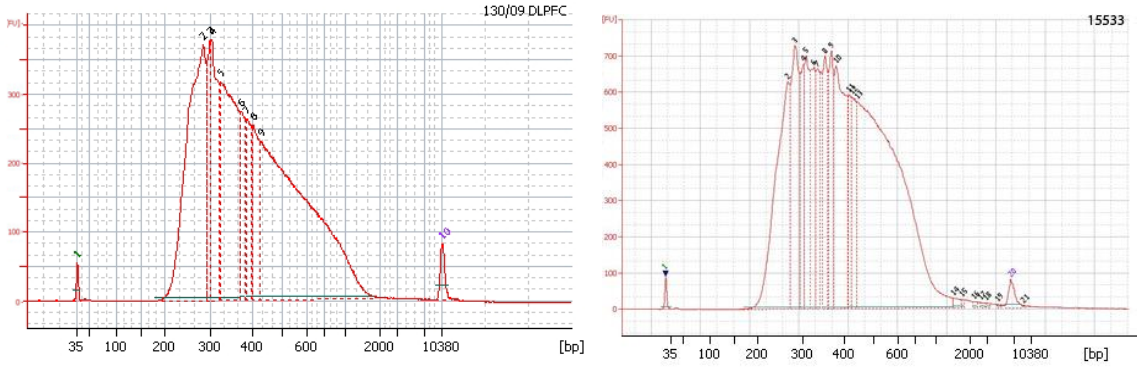


Figure 2. Capillary electrophoresis profiling of the RNA-seq libraries. Adult DLPFC (left) and whole foetal brain (right) samples were analysed in the Agilent High Sensitivity DNA chip using the Agilent 2100 Bioanalyzer. X-axis indicates the size of DNA fragments, Y-axis represents the fluorescence intensity (FU).

3.3.5. Bioinformatic analyses

3.3.5.1. Metrics and Quality Control

Quality of the sequencing was initially assessed by FastQC (Appendix 2) (Andrews, 2010). Metrics for these samples were compatible with sequencing of RNA. High depth sequencing of libraries from DLPFC and foetal tissue generated 167,943,560 and 153,548,265 reads, respectively. Raw reads were trimmed using

Trimmomatic v0.33 (Bolger et al., 2014) for removal of leading and trailing sequences with a phred-33 score below 15 (which means that allocated bases with calculated accuracy <97% were excluded), and for removal of Illumina universal adapters. Quality of the sequences was further scanned by a sliding window of four nucleotides to exclude sequences with an average phred score <15, and only reads with a minimum of 36 bases were kept. The number of reads mapped to the genome were 156,892,943 for the DLPFC sample and 148,043,166 for the human foetal brain sample, representing a removal of 6.6% (11,050,617 reads) and 3.6% (5,505,099 reads) of poor sequences, respectively.

3.3.5.2. Mapping reads and assembling transcripts

Trimmed reads were aligned to the annotated reference human genome hg19 using TopHat v2.1.0 (Kim et al., 2013). Overall mapping of reads to the annotated reference genome were of 95.5% for the DLPFC sample and 94.7% for the foetal brain sample. TopHat was run using the average maximum inner distance between mate pairs as 500 bases (-r 500). Cufflinks (Trapnell et al., 2010) was used to assemble reads into transfrags using the genome build hg19 as reference. Cuffcompare (Trapnell et al., 2010) was used to mount transfrags into transcripts and to estimate transcript abundance by assigning FPKM values to each transcript. FPKM refers to "fragments per kilobase of exon per million fragments mapped" and is a unit of gene expression in RNA-seq analysis. Although units of FPKM are provided in the Results section of this chapter, it is important to refrain from comparing these values between samples, as this project is underpowered. Transcripts for which FPKM values were found within the first quartile of the RNA-seq data were considered not expressed. Junction reads were observed on the Integrative Genomics Viewer v2.3 (www.broadinstitute.org/igv/).

3.3.6. Reverse transcription

RNA samples from independent subjects were used to validate potential novel transcripts identified by RNA-seq. DNA-free RNA samples were reverse transcribed, PCR-amplified and analysed on agarose gels. Reverse transcription was performed as in Chapter 2. Briefly, DNase-treated RNA samples were converted to cDNA by using approximately one µg DNA-free RNA and SuperScript III® Reverse Transcriptase. The two-step reaction led to a final 20 µL volume reaction containing 5 µM random decamers, 500 µM dNTPs, 5 mM DTT, 40 units RNaseOUT™ Recombinant Ribonuclease Inhibitor, 200 units of SSIII reverse transcriptase enzyme in 1x First-Strand Buffer and approximately one µg RNA. The initial mix containing RNA, random decamers and dNTPs was heated at 65ºC for 5 min and placed on ice for at least 1 min, when the remaining components were added. The reaction was incubated in a GS4 Thermocycler (G-Storm, Somerset, UK) with a heated lid for 25 min at 25ºC, 60 min at 50ºC, 30 min at 55ºC and 15 min at 70ºC. cDNA was diluted 1:7 prior to use and stored at -20ºC. Reagents were purchased from ThermoFisher Scientific.

3.3.7. Reverse transcription PCR (RT-PCR)

Expression of the RefSeq transcripts of *BORCS7* and *NT5C2*, as well as the novel transcripts of *NT5C2* observed in the RNA-seq data, was qualitatively investigated in different samples of brain regions using RT-PCR. Transcript-specific oligonucleotides were designed using Primer3 (Untergasser et al., 2012), and these were purchased from Integrated DNA Technologies (IDT, London, UK) (Table 2). PCR amplification of cDNA was performed using the HOT FIREPol® DNA Polymerase (Solis Biodyne, Tartu, Estonia) in 12 µl reactions. These consisted of 6 µL cDNA (5-50 ng) with 1× HOT FIREPol Buffer B1, 2.1 mM MgCl2, 200 µM dNTP set and 0.5 unit of HOT FIREPol Taq Polymerase. The reaction was incubated for 15 min at 95ºC (hot start), 35 cycles of 95ºC for 30 sec, 60ºC for 30 sec, 72ºC for 30 sec, and a final elongation step of 10 min at 72ºC.

Table 2. Oligonucleotides used for PCR-amplification.

Gene	Transcript (RefSeq)	Forward oligonucleotides 5'-3'	Reverse oligonucleotides 5'-3'	Expected size in cDNA (bp)
BORCS7	NM_001136200	AATGACATGTAAGAGTGCTGTAGGAC	CCGGCCCTAACACAGAACTTT	168
	NM_144591	ACAACACATCTTCAATACCAGCAAG	CGTATCTCCTCTGAGTTAATGTCATTC	119
NT5C2	NM_012229	CCGAGGCGAATGGATCACTTG	GCGCTGGAGCCGAGTTTC	84
	NM_001134373	CATATCTGCTGCATTCTGTAACCGA	GTGCGCTGGAGCCGAATAC	83
	d3 (new)	CAAAGCTGAGCAACTCCTGG	CGAGAAGCCTATCATCGTGTAC	110
	e3.1 (new)	TCCACAGTAAGCTCAAAACCAA	ATACCCTTGCTGGAGAGCC	122
	e3.1 d3 (new)	GCTCAAAACCAAGGGACTCA	AGAAGCCTATCATCGGAGAGC	115
SDHA	All transcripts	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCATG	86

3.3.8. DNA gels and visualisation

PCR products were visualised in ethidium bromide-stained agarose gels, which consisted of 2.5% (w/v) agarose (Apollo Scientific, Cheshire, UK) in 1x TAE buffer (40 mM Tris, 20 mM acetic acid, and 1 mM EDTA). Ethidium bromide 10 mg/mL 0.001% (v/v) was added, and the solution was cooled in an electrophoresis bed for 40 min at room temperature. The gel was immersed in electrophoretic buffer 1X TAE, and aliquots of RT-PCR product were loaded with Orange G-based loading buffer (final concentration: 17 mg Orange G in 5% glycerol). Electrophoresis was run at 120 V for approximately 35 min. Gels were imaged in a GelDoc-It® Imager (UVP, Cambridge, UK) and recorded as TIF files.

3.3.9. Other in silico analyses

Expression of chromosome 10q24 genes was investigated in the public database Human Brain Transcriptome (http://hbatlas.org/) (Kang et al., 2011), which contains whole genome microarray expression data from different brain regions in different time points of development. Unpublished data obtained by collaborators using Illumina HumanHT-12 v4 Expression BeadChips were used to investigate the expression of *BORCS7, AS3MT, CNNM2,* and *NT5C2* during *in vitro* neuronal differentiation. These studies examined the global gene expression profile of the cortical neural progenitor cells CTX0E16 submitted to a 28-day differentiation protocol (differentiation protocol detailed in Anderson et al. (2015)), and of hippocampal neural progenitor cells HPC03A/07 submitted to a 7-day differentiation

protocol (Powell et al. (2016), submitted). Differentially expressed genes were considered as those for which probe intensity significantly differed from the neural progenitor stage, at a false discovery rate cut-off (FDR)<0.05.

3.4. Results

3.4.1. Transcript discovery and microarray expression data

Schizophrenia risk variants on chromosome 10q24 were found to be associated with altered *cis*-regulation of *BORCS7, AS3MT* and *NT5C2* in Chapter 2. The pilot RNA-seq study presented in this chapter aimed at identifying the transcripts of these genes produced in brain regions where the putative risk mechanisms were observed. Briefly, all RefSeq transcripts of these genes were detected in two brain tissues, with the exception of the read-through *BORCS7-AS3MT* transcript (Table 3). The Tuxedo pipeline did not detect any novel transcript variants, but the manual inspection of exon junction counts in the RNA-seq data provided evidence to suggest the existence of less expressed transcripts of *AS3MT* and *NT5C2*.

BORCS7 transcript NM_001136200 was found more expressed in the adult DLPFC (8x) and the foetal brain (nearly 100x) than NM_144591. These transcripts differ in the 3' UTR (untranslated region), which could interfere with transcript decay or localisation. No evidence for novel transcripts was found in the RNA-seq data. Expression of *BORCS7* in several brain regions was observed to peak at about one year after birth and to gradually decrease throughout adulthood (Figure 3A) (Kang et al., 2011). Similarly, expression of this gene was found to increase during neuronalisation of our in-house cortical (fold-change=1.46; FDR=0.01) and hippocampal cellular models (fold-change=1.12; FDR=3.22E-04). These data suggest that *BORCS7* could be involved in early neuronal specialisation.

Table 3. Expression of chromosome 10q24 genes in the RNA-seq data.

Genes	Transcript	FPKM*	
		Adult DLPFC	Foetal brain
BORCS7	NM_001136200	2.35	3.95
	NM_144591	0.29	0.04
BORCS7-AS3MT	NR_037644	-	-
AS3MT	NM_020682	1.42	4.79
CNNM2	NM_199077	0.14	0.13
	NM_017649	1.55	0.85
	NM_199076	0.86	0.93
NT5C2	NM_012229	1.41	3.57
	NM_001134373	6.43	9.76

*FPKM = fragments per kilobase of transcript per million mapped reads. The cut-off value used to indicate expression of a given gene varies according to library size and sequencing depth. All FPKM values shown in this table indicate expression of all transcripts, which have been validated by RT-PCR on Figure 6.

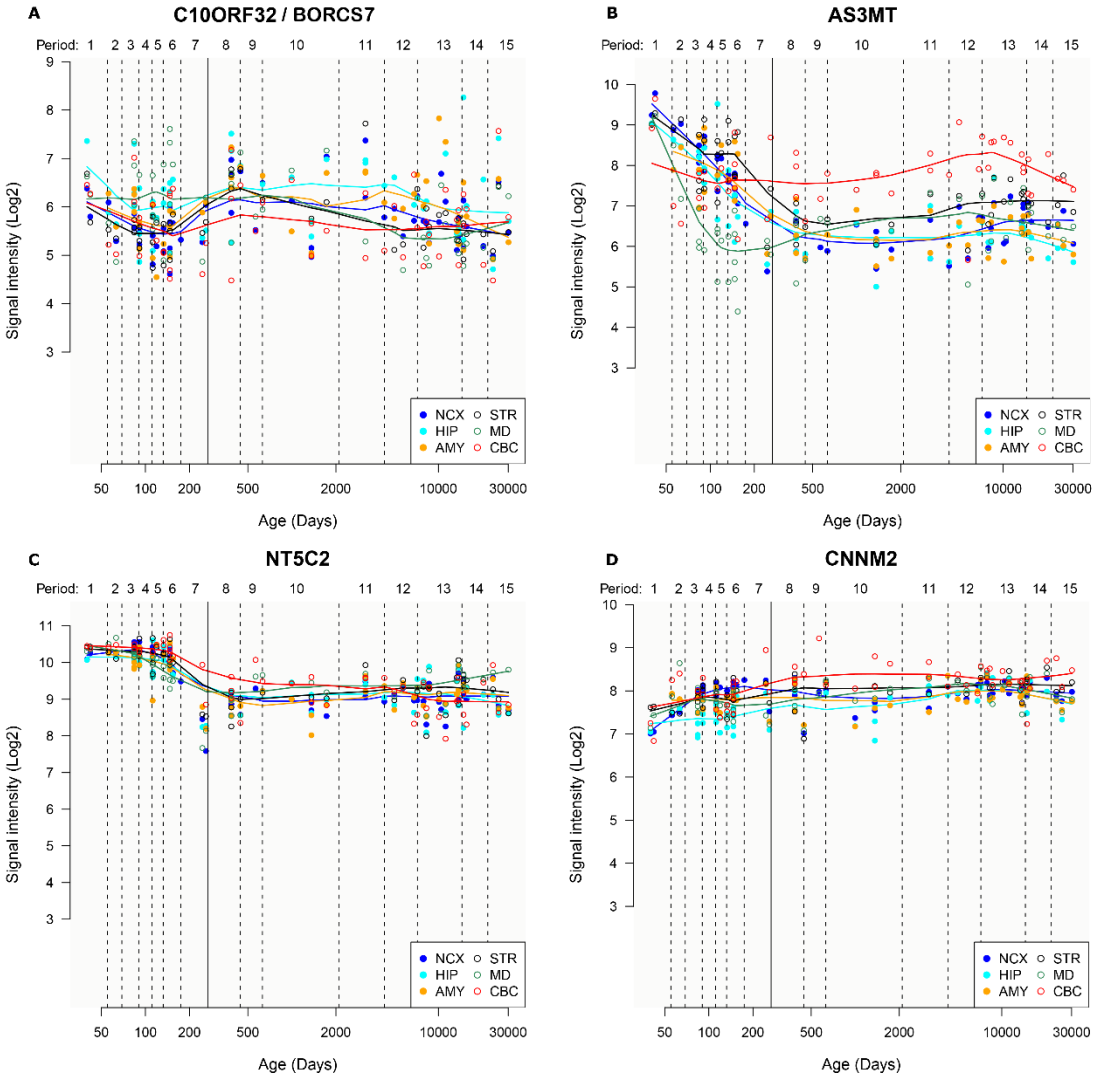


Figure 3. Expression of chromosome 10q24 genes in specific brain regions across life, according to microarray data from Kang et al. (2011). Signal intensity (log2) is given on the Y-axis, and age (days since conception) is given on the X-axis. **A**, BORCS7; **B**, AS3MT; **C**, NT5C2; **D**, CNNM2. Legend: neocortex (NCX), hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of thalamus (MD), cerebral cortex (CBC).

As for *AS3MT,* only the expression of NM_020682 was significantly detected in the RNA-seq data analysis. The manual inspection of junction reads, however, provided evidence for the existence of *AS3MT*^{d2d3}*,* proposed by Li et al. (2016) as the putative schizophrenia susceptibility transcript (blue arrows*,* Figure 4). It is possible to extrapolate from the RNA-seq data presented in this chapter, based on counts of junction reads, that the full length transcript of *AS3MT* is more abundantly expressed than *AS3MT*^{d2d3}. Exon junction counts for transfrags associated with *AS3MT*^{d2d3} transcript (<5 counts) were sporadic when compared to the number of transfrags associated with full length *AS3MT* (approximately 200 counts) in the foetal brain sample, for instance. An accurate estimate of the relative expression of these transcripts, however, would require a more sophisticated analysis and the inclusion of additional samples. Nonetheless, overall expression of *AS3MT* was found to be maximal during foetal development, as opposed to adulthood, according to public microarray data (Figure 3B). Consistent with this*, AS3MT* expression in differentiated hippocampal neurons was significantly lower than at their neural progenitor stage (fold-change=0.87, FDR=4.52E-14), although this was not observed in the cortical cell line (fold-change=1.03, FDR=0.47). It is important to note that the microarray probe tagging *AS3MT* expression in these microarray studies (Illumina HT12 v4 BeadChip) is specific to the most 3' exon of this gene, and therefore it would not distinguish between expression from full length *AS3MT* or *AS3MT*^{d2d3}. Nonetheless, these data suggest a potential role for *AS3MT* in early neural development.

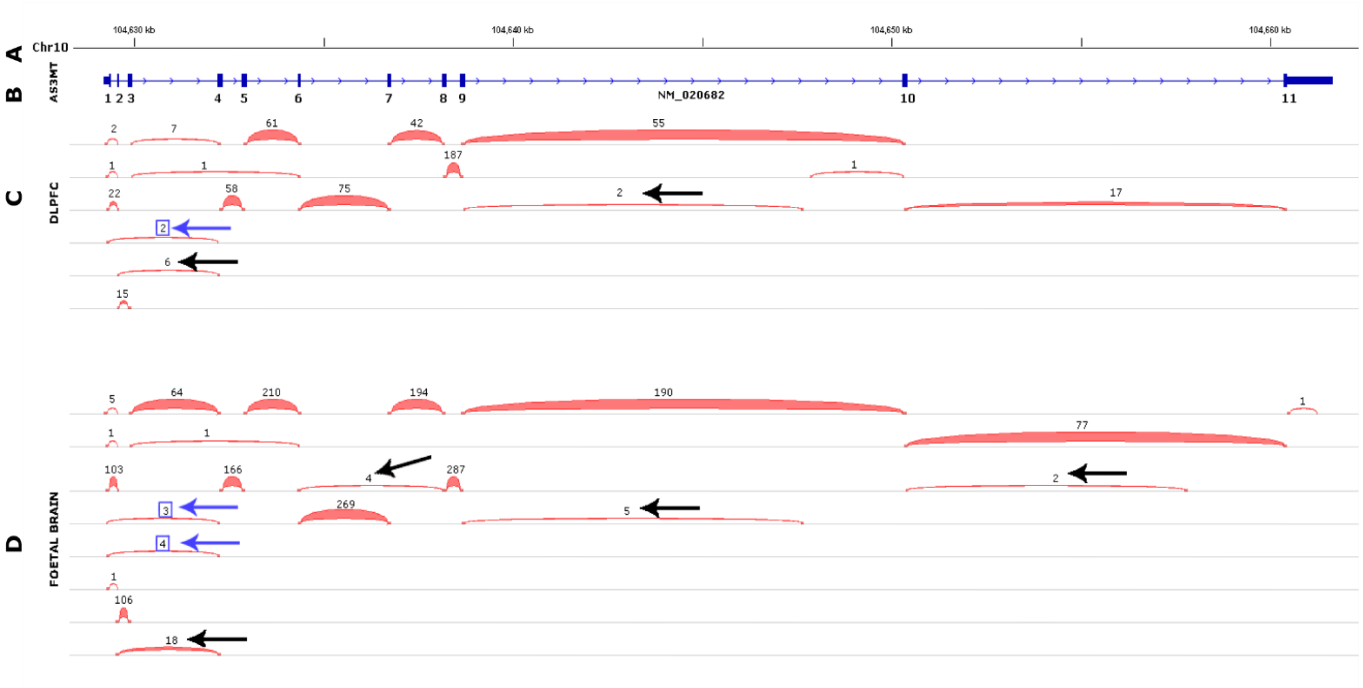


Figure 4. Counts of junction reads observed for *AS3MT* in RNA-seq data. **(A)** Position of *AS3MT* on chromosome 10. **(B)** The full length RefSeq transcript of *AS3MT* found using Cufflinks and Cuffmerge. **(C, D)** Counts of exon junctions in the DLPFC sample and in the foetal brain in the RNA-seq data for *AS3MT*. This analysis indicates the presence of a transcript lacking exons 2 and 3 (blue arrows), which would indicate the presence of AS3MT^{d2d3}, suggested by Li et al. (2016) as the schizophrenia susceptibility gene. Additional transcripts can be inferred by certain transfrags (black arrows), although they are very lowly expressed. An increased number of samples would provide statistical power in order to obtain an accurate measurement of expression of these transcripts.

The expression of the NT5C2 transcript NM_001134373 was found to be higher than its counterpart, transcript NM_012229, in the DLPFC (4x) and foetal brain (3x). Both *NT5C2* transcripts seem to encode identical proteins, although NM_012229 has an additional 5' UTR region, which may indicate an additional level of regulation. Exon junction counts observed in the RNA-seq data (Figure 5) suggest the existence of a relatively common new exon between exons 3 and 4 of NM_001134373, or exons 4 and 5 of NM_012229, which is simply denoted here as exon 'exon 3.1'. Moreover, it was observed the existence of transfrags associated with transcripts in which exon 3 of NM_001134373, or exon 4 of NM_012229, referred here simply as 'exon 3', was absent. These data suggest the existence of novel transcripts of *NT5C2*: *NT5C2*$^{e3.1}$ (contains exon 3.1), *NT5C2*^{d3} (lacks exon 3), and *NT5C2*$^{e3.1\ d3}$ (contains exon 3.1 and lacks exon 3). Based on junction reads counts, it appears that the full length transcripts of *NT5C2* are more expressed than these novel variants. In the RNA-seq data from foetal brain, for example, junction count reads corresponding to transfrags of full length *NT5C2 (*approximately 300 counts*)* were more than 4x more abundant than transfrags associated with *NT5C2*$^{e3.1}$ (around 70 counts), which in turn corresponded to almost double the number of junction reads associated with *NT5C2*^{d3} and *NT5C2*$^{e3.1\ d3}$ (approximately 30 counts each). The unbiased quantification of expression of these transcripts, however, would require the inclusion of more samples, and validation using rapid amplification of cDNA ends (RACE) or qPCR, for example. Overall *NT5C2* expression was found to be maximal during foetal development rather than in adulthood in several brain areas (Figure 3C), and was similarly increased in differentiating hippocampal neurons, as compared to their neural progenitor stage (fold-change=1.40, FDR=4.38E-20). No statistically significant change in expression of *NT5C2* was observed for the cellular model of cortical neuronal differentiation CTX0E16 (fold-change=1.09, FDR=0.20), which possibly suggests that the gene might be involved in localised neuronal differentiation.

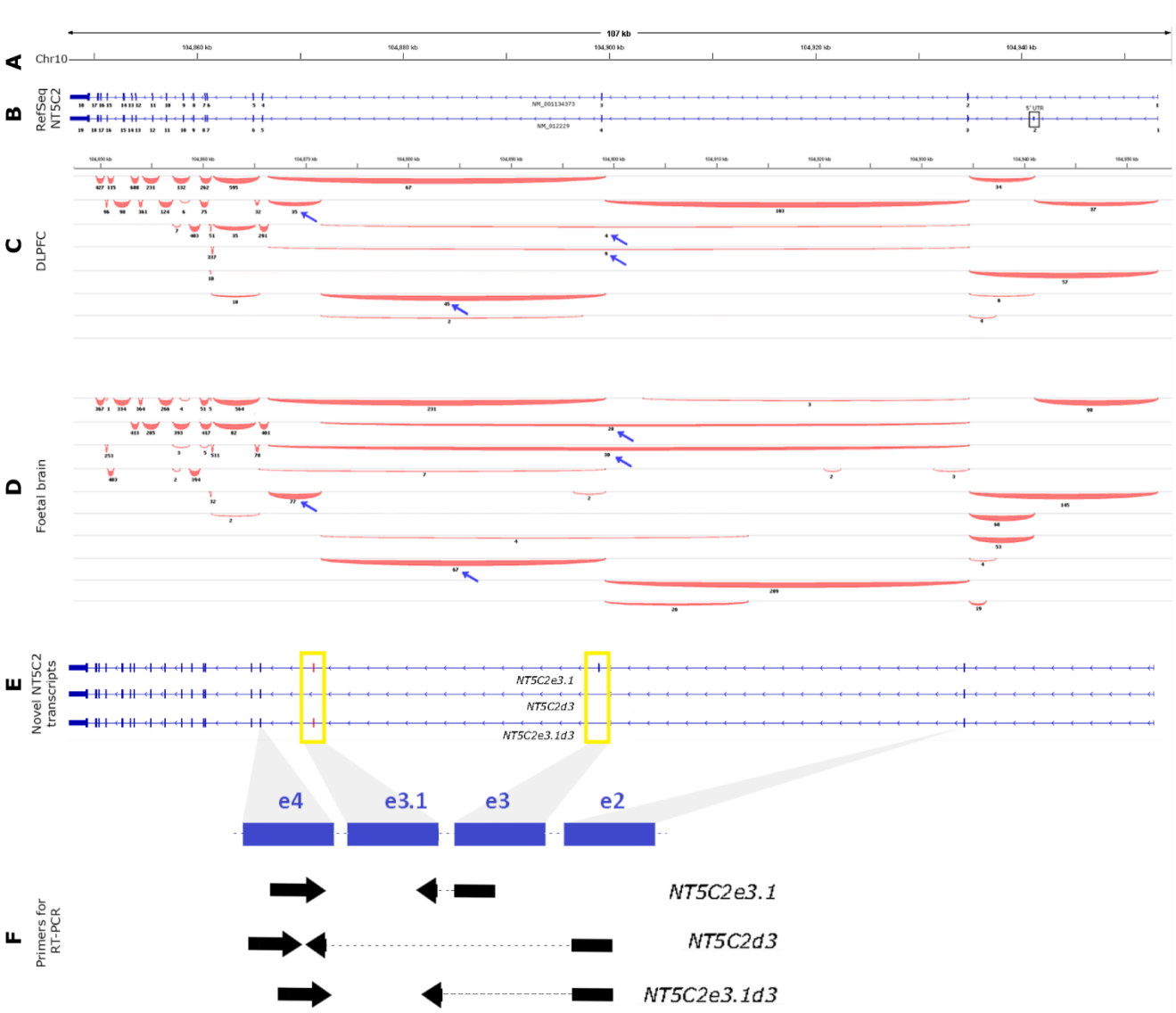


Figure 5. Counts of junction reads observed for NT5C2 in the RNA-seq data. **(A)** Location of this gene on chromosome 10. **(B)** RefSeq transcripts of *NT5C2* found by Cufflinks and Cuffmerge in the two samples. **(C, D)** Counts of exon junctions for *NT5C2* in the DLPFC sample and in the foetal brain. Blue arrows indicate junction reads that appeared in both tissues. An increased number of samples would provide statistical power in order to obtain an accurate measurement of expression of these transcripts. **(E)** Novel putative *NT5C2* transcripts. Yellow boxes indicate main differences found. **(F)** Representation of binding location sites for primers designed for RT-PCR to flank exon-exon junctions, in order to amplify transcript-specific amplicons. Blue dashed lines indicate intronic sequences, black arrows indicate primer sequences.

Expression levels of *CNNM2* transcripts in the RNA-seq data is also reported (Table 3), although schizophrenia risk variation was not associated with cis-regulatory mechanisms on this gene (e.g. Chapter 2, Li et al., 2016). Expression of this gene was not observed to vary during development (Figure 3D) (Kang et al., 2011) or during neuronalisation of the cortical (fold-change=0.92, FDR=0.06) or hippocampal in-house cellular models (fold-change=1.00, FDR=0.35).

3.4.2. Validation of transcripts in other samples of different brain areas

Expression of the RefSeq transcripts of *BORCS7 (*NM_001136200 and NM_144591) and *NT5C2* (NM_012229 and NM_001134373), as well as of the newly identified transcripts $NT5C2^{d3}$, $NT5C2^{e3.1}$ and $NT5C2^{e3.1\ d3}$, and of the housekeeper *SDHA* (succinate dehydrogenase complex, subunit A) was investigated in four samples each of foetal brain, and adult DLPFC, hippocampus and caudate (Figure 6). Samples analysed by RNA-seq were also included, represented by lanes #1 (DLPFC) and #13 (foetal brain). *BORCS7* RefSeq transcript NM_1136200 was found to be expressed in all samples, although transcript NM_0144591 was not detected in some samples of DLPFC and hippocampus. Both *NT5C2* RefSeq transcripts were detected in all samples, but as for the novel transcripts, only $NT5C2^{e3.1}$ was found expressed in all samples, whereas $NT5C2^{e3.1\ d3}$ and $NT5C2^{d3}$ were not. This is consistent with the observation of exon junction counts in the RNA-seq data, as junction counts associated with exon e3.1 were more abundant than junction counts associated with transcripts lacking exon 3 ($NT5C2^{e3.1\ d3}$ and $NT5C2^{d3}$). Additional expression studies are warranted to investigate possible transcript-specific mechanisms driving risk for schizophrenia on chromosome 10q24, especially regarding *NT5C2*. Nonetheless, full length transcripts for all genes were observed as the most expressed messages.

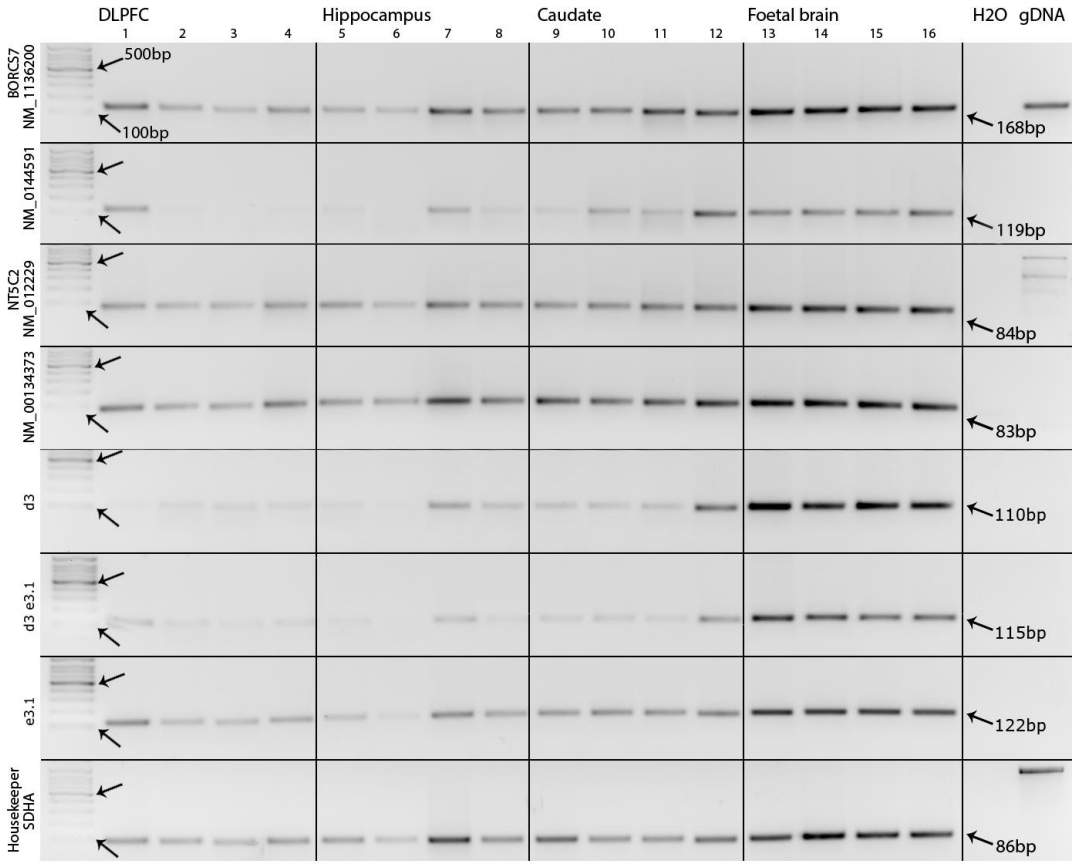


Figure 6. Validation of RNA-seq findings by RT-PCR in different brain tissues. Expression of RefSeq transcripts NM_1136200 and NM_0144591 of *BORCS7,* and NM_012229 and NM_00134373 of *NT5C2*, alongside with the novel *NT5C2* transcripts and a housekeeper gene (SDHA) was investigated in 16 brain samples. Arrows on the ladder column indicate the marks of 500bp and 100 bp. The right column indicates the expected size for each amplicon. Amplification with genomic DNA is also shown, as primers for NM_1136200 and *SDHA* were designed to amplify using cDNA or gDNA, while other primer sets were designed to span exon-exon junctions and thus are cDNA specific.

3.5. Discussion

The collection of RNA messages expressed at a specific time point or location may inform of biological mechanisms related to disease (Wang et al., 2009). In this chapter, RNA-sequencing was used to investigate chromosome 10q24 transcripts expressed in brain tissues where *cis*-regulatory effects associated with schizophrenia risk alleles were observed in Chapter 2. Annotated transcripts of genes putatively implicated in susceptibility to schizophrenia, *BORCS7, AS3MT* and *NT5C2*, as well as novel transcript variants of *NT5C2* were detected in different brain areas implicated in schizophrenia, including the foetal brain and the adult DLPFC, hippocampus and caudate nucleus. Evidence to support the existence of $AS3MT^{d2d3}$ in the adult DLPFC and in the second trimester foetal brain was also observed.

This chapter arose from the interesting observations that genome-wide association signals may index *cis*-regulatory variants that affect specific gene transcripts (Li et al., 2016; Tao et al., 2014). The non-coding SNP rs1344704, for example, the first to exhibit genome-wide association with psychosis (O'Donovan et al., 2008) and schizophrenia (Riley et al., 2010; Schwab et al., 2013; Williams et al., 2011; Zhang et al., 2011; Zhang et al., 2012), was associated with *cis*-regulatory effects on *ZNF804A* exclusively in the developing foetal brain (Hill and Bray, 2012). More recently, however, the *cis*-regulatory risk mechanism was suggested to alter expression of a specific transcript, termed $ZNF804A^{E3E4}$ (Tao et al., 2014). As for chromosome 10q24, genotype at rs7085104 (which is in strong LD with rs11191419: r^2=0.79) has been recently suggested to elicit *cis*-regulatory effects on a specific transcript of *AS3MT*, $AS3MT^{d2d3}$ (Li et al., 2016). Interestingly, the full length transcript of *AS3MT* encodes a methyltransferase enzyme involved in arsenic metabolism (Sumi and Himeno, 2012), but $AS3MT^{d2d3}$ does not encode the methyltransferase domain, and therefore its function remains unknown. This chapter therefore intended to determine the existence of potentially novel transcripts encoded in chromosome 10q24.

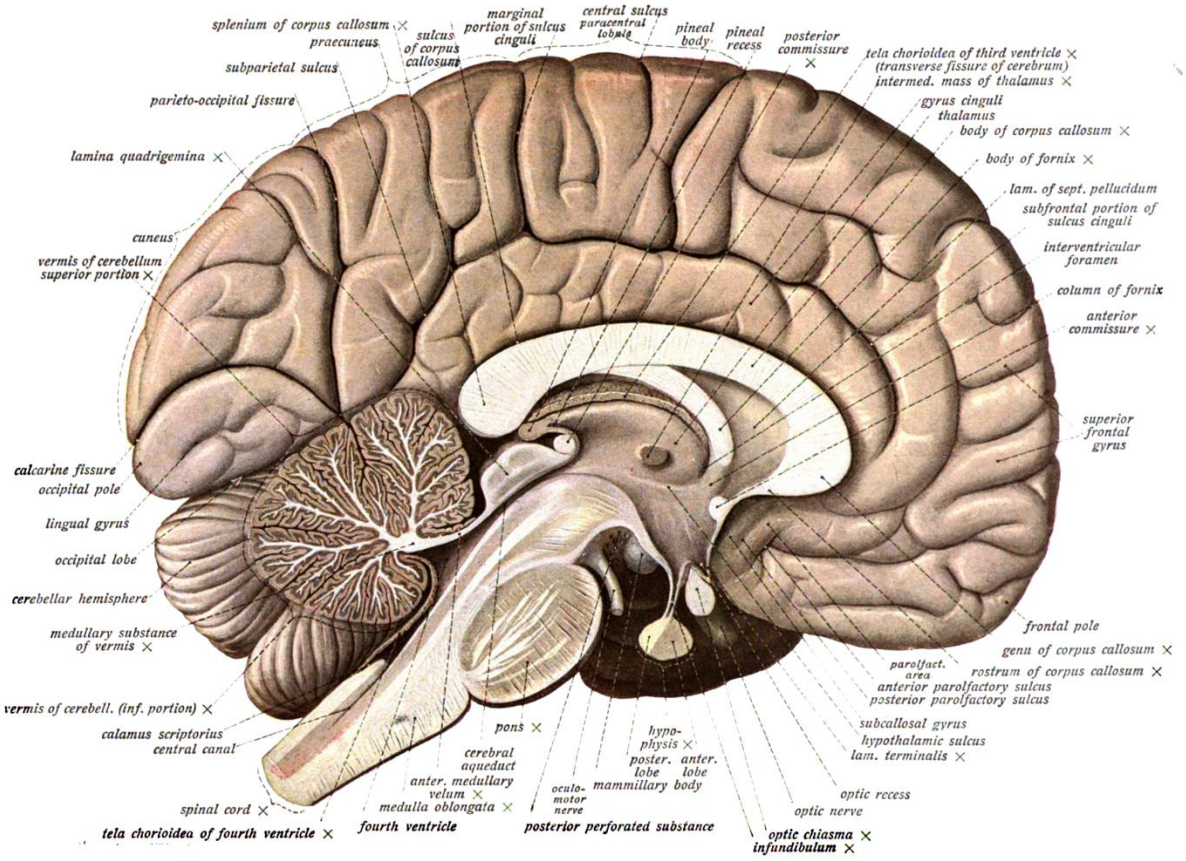
The small number of samples included in the RNA-seq analysis precluded the confident identification of any transcripts of *AS3MT* other than its full length variant. However, manual inspection of junction reads in the RNA-seq data provided evidence to support the expression of $AS3MT^{d2d3}$ in the adult and foetal brains. The expression of the truncated version has been previously estimated to be approximately five-fold lower than the expression of full length *AS3MT* in healthy subjects (Li et al., 2016). This would require that *cis*-regulatory effects specific to this novel transcript to be very large, in order to result in the effects on allelic expression of *AS3MT* observed in Chapter 2, in which both transcripts would have been indexed. While it is out of the scope of this pilot study to provide accurate measures of expression of specific transcripts, we found evidence which suggests that the expression of $AS3MT^{d2d3}$ is much lower than that of *AS3MT*. Overall *AS3MT* expression was found to be decreased in several brain regions during adult life, and in the *in vitro* model of hippocampal neuronalisation, providing evidence to support the role of this gene in neurodevelopment.

No evidence in the RNA-seq data was found to support the existence of novel transcripts of *BORCS7,* or the read-through *BORCS7-AS3MT*, in agreement with findings from Li et al. (2016). Expression of full length BORCS7 transcripts NM_001136200 and NM_144591, however, was confirmed in multiple foetal and adult brain samples. Moreover, *BORCS7* expression was found to significantly increase during hippocampal and cortical neuronalisation, as well as after birth, and to gradually decrease throughout adulthood. The gene encodes a subunit of the BLOC-1 related complex, which is implicated in lysosomal function, cell migration and neurite outgrowth (Ghiani et al., 2010; Pu et al., 2015). The association of this gene's allelic expression with schizophrenia risk variants may suggest that these processes may be affected in disease or disease predisposition mechanisms.

NT5C2 encodes the *5'*-cytosolic nucleotidase II, which is involved in the regulation of intracellular nucleotide and nucleoside pools (Bianchi and Spychala, 2003), and therefore is relevant to the purinergic hypothesis of schizophrenia (Lara

and Souza, 2000). Overall *NT5C2* expression is maximal during foetal development in several brain areas, and was found to be increased during hippocampal neuronalisation, supporting the role of *NT5C2* in neurodevelopment. Expression of three novel transcript variants of *NT5C2* was confirmed in different brain tissues (Figure 5). Some evidence was observed to suggest that full length *NT5C2* is more abundant than its truncated counterparts, although the precise quantification of their expression would require a more sophisticated analysis and the inclusion of more samples. It is possible that the novel transcript variants may still encode the nucleotidase, as the nucleotidase domain is encoded further downstream from exons 3 and 3.1, but more studies are warranted to elucidate their function.

In summary, this chapter provided evidence which suggests that RefSeq transcripts of *BORCS7, AS3MT* and *NT5C2,* as well as the novel transcripts $AS3MT^{d2d3}$, $NT5C2^{e3.1}$, $NT5C2^{e3.1\ d3}$ and $NT5C2^{d3}$, are present in multiple samples of brain tissue. They may be responsible for driving the cis-regulatory effects observed as putative schizophrenia risk mechanisms in Chapter 2. However, the expression of full length *AS3MT* and *NT5C2* transcripts are much higher than the novel variants. This would require *cis*-regulatory effects specific to the truncated transcripts to be very large, in order to result in the effects on allelic expression observed for these genes in Chapter 2, where all their transcripts would have been indexed. An analysis including more samples is warranted to further elucidate the putative risk mechanisms on chromosome 10q24 driving association with schizophrenia.



Coronal view of the human brain. Illustration from the Sobotta's Human Anatomy Atlas (1908).

Chapter 4

Distribution of NT5C2 protein in the adult human brain and in human cellular models of neurodevelopment

4.1. Abstract

Multiple *cis*-regulatory effects on chromosome 10q24 were detected in the adult and foetal brain in association with risk for schizophrenia, in Chapter 2. Of these, the decreased expression of *NT5C2* in association with both schizophrenia risk alleles in the dorsolateral prefrontal cortex (DLPFC) was a prominent finding. This gene encodes the cytosolic 5'-nucleotidase II (NT5C2), a homotetramer hydrolase involved in the regulation of intracellular nucleotide and nucleoside pools. While *NT5C2* is conserved from Prokaryotes to Eukaryotes with various degrees of sequence identity, determining the neural cell type(s) which express the enzyme in the human brain, and its distribution in neural progenitor cells (NPCs) may reveal clues of the physiological roles relevant to schizophrenia, associated with this enzyme. The objective of this chapter, therefore, is to examine the distribution of NT5C2 in the adult DLPFC and human neural stem cells. For this, NT5C2 expression was investigated in human NPCs derived from the foetal brain and from human induced pluripotent stem cells (hiPSC). The investigation was performed using fluorescence immunocytochemistry staining and visualisation under confocal microscopy, which revealed NT5C2 distribution spread throughout the soma and cellular processes of the NPCs. In the *post-mortem* adult DLPFC, NT5C2 expression was investigated by means of immunohistochemistry staining using 3,3'-diaminobenzidine (DAB) and Nissl. The enzyme was detected in neurons, glial cells and the neuropil, although glial cells for which NT5C2 staining was absent were also observed. These findings are in agreement with a ubiquitous occurrence of this enzyme given its widespread presence in the brain and NPCs, but cannot exclude a possible cell type-specific schizophrenia risk mechanism.

Note: The neural progenitors derived from human induced pluripotent stem cells (hiPSCs) were kindly provided by Dr Carole Shum (Srivastava's lab) in collaboration with Prof Price's group. Immunohistochemistry was advised by Dr Marie-Caroline Côtel (Vernon's lab) and Dr Claire Troakes (MRC Brain Bank). Images were obtained with the assistance of Dr Hemanth Nelvaga and Dr Marta Tarczyluk (Cooper's lab).

4.2. Introduction

The cytosolic 5'-nucleotidase II (NT5C2, or cN-II) is a widely expressed homotetrameric enzyme that consists of 561 amino acids. This enzyme catalyses the dephosphorylation of purine monophosphates, mainly inosine and guanosine monophosphate (IMP and GMP, respectively), and the transfer of phosphate groups between nucleotides and nucleosides (Cividini et al., 2015a). NT5C2 is critically positioned at the crossroad between *de novo* nucleotide synthesis and salvage pathways, acting as a key regulator for these processes and therefore controlling intracellular nucleotide and nucleoside levels (Hunsucker et al., 2005). NT5C2 is part of the family of 5'-nucleotidases, together with a membrane-bound extracellular enzyme (eN), a mitochondrial nucleotidase (mdN) and four other cytosolic enzymes (NT5C1A, NT5C1B, NT5C3 and NT5C3L) (Itoh, 2013). Their difference, besides subcellular localisation, relies on tissue-specificity and altered affinities for derivatives of nucleotide monophosphates (NMPs, such as IMP, GMP, AMP, etc.). These characteristics allow this family of enzymes to fine tune intracellular signalling molecules according to the metabolic needs of a particular tissue or cell type (Hunsucker et al., 2005).

NT5C2 is encoded on chromosome 10q24, a genomic region robustly implicated in schizophrenia by large-scale genome-wide association studies (Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Schizophrenia Working Group of the PGC, 2014). In Chapter 2, decreased expression of *NT5C2* in different brain tissues was associated with schizophrenia risk genotype, an event that was very noticeable in the adult DLPFC. In this brain region, the majority of *cis*-regulatory effects observed on *NT5C2* could be accounted for by genotype at the top two risk variants at the locus, rs11191419 and chr10_104957618_I. Interestingly, the risk indel chr10_104957618_I is in linkage disequilibrium with rs12413409 and rs11191548 (r^2=0.82 in both cases), genome-wide significant association signals for body mass index and cardiac disease (Newton-Cheh et al., 2009; Schunkert et al., 2011), suggesting that variation in the gene might have pleiotropic effects.

The distribution of NT5C2 in the nervous system, however, is poorly characterised. The objective of this chapter is to investigate the expression of this enzyme in the adult dorsolateral prefrontal cortex, a region that has been implicated in schizophrenia pathophysiology (Callicott et al., 2000; Galindo et al., 2016), and additionally in NPCs derived from the foetal brain and from hiPSCs, which recapitulate early stages of neural development. As expression of *NT5C2* in the cortex has been described in the Human Protein Atlas (Figure 1) (Uhlen et al., 2015), this chapter focused on replicating these findings specifically in the DLPFC. Moreover, several human cellular models will be used in this chapter to explore immunostaining methods (western blotting, immunohistochemistry and immunocytochemistry), what will be important for the next chapter. These cell models include the human embryonic kidney cells HEK293T, the bone marrow SH-SY5Y cells, the NPCs derived from human induced pluripotent stem cells (hiPSCs) (Cocks et al., 2014; Shi et al., 2012) and from the neuroepithelium of a second trimester foetal brain, the CTX0E16 cell line (Anderson et al., 2015; Pollock et al., 2006). This investigation may implicate specific cell types associated with NT5C2 expression, which could be giving rise to the putative risk mechanism associated with schizophrenia in Chapter 2.

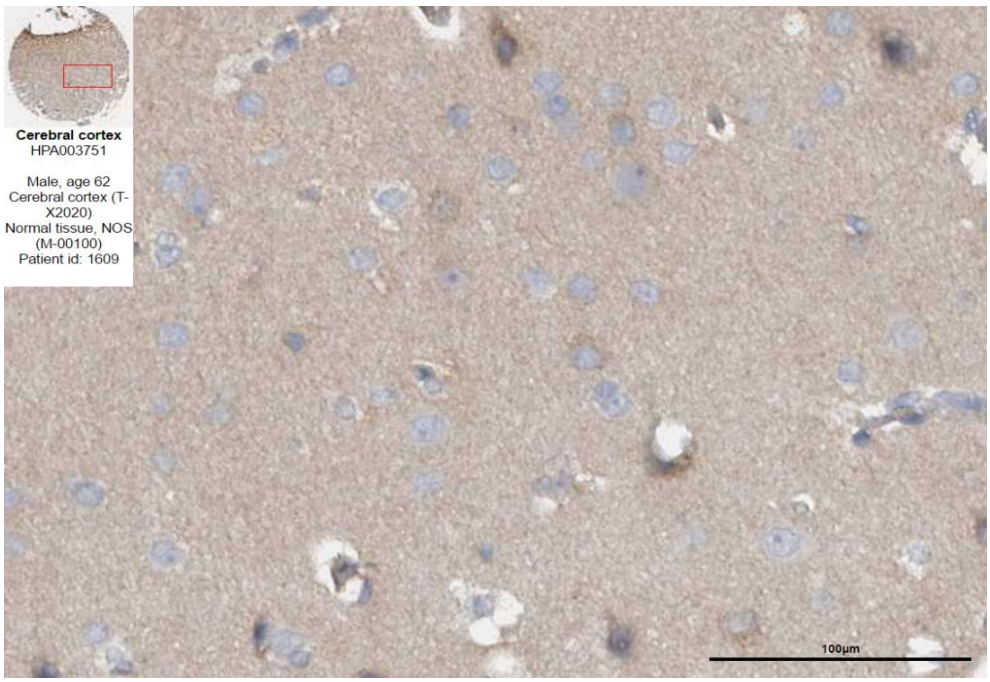


Figure 1. NT5C2 immunohistochemistry of the cerebral cortex extracted from the Human Protein Atlas (http://www.proteinatlas.org/) (Uhlen et al., 2015) using the mouse monoclonal NT5C2 antibody (3C1). Quantification of cells positive for NT5C2 in this database showed expression of NT5C2 in over 75% of neurons and of the neuropil, with glial cells found to stain in 25-75% of accounted instances.

4.3. Methods

4.3.1. Cell lines

Antibodies used to detect NT5C2 protein were validated and tested in (1) the human embryonic kidney HEK293T cells, (2) the bone marrow SH-SY5Y cells, (3) the conditionally-immortalised cortical neural progenitor cells (NPCs) CTX0E16 and (4) the cortical NPCs derived from human induced pluripotent stem cells (hiPSC) from a healthy individual. All cells lines were grown in Nunclon™ Delta Surface tissue culture flasks or plates (ThermoFisher Scientific, Waltham, MA, USA). HEK293T and SH-SY5Y cells were maintained at 37ºC with 5% CO_2, in T75 flasks containing media that consisted of Advanced Dulbecco's Modified Eagle Medium/Ham's F-12 (Advanced DMEM/F-12) (ThermoFisher Scientific, Waltham, MA, USA) supplemented with 10% (v/v) Foetal Bovine Serum (ThermoFisher Scientific) and 2 mM GlutaMAX™ Supplement (ThermoFisher Scientific), fed every 2-3 days, and passaged using Accutase (Sigma, St. Louis, MO, USA) once 80–90% confluency was achieved. The conditionally immortalised neural stem cells CTX0E16 cells (Anderson et al., 2015; Pollock et al., 2006), which are further discussed in Chapter 5, were obtained from ReNeuron (www.reneuron.com) as part of a Material Transfer Agreement. These cells were maintained at 37ºC and 5% CO_2, in T75 flasks previously coated with 1 µg/cm^2 Engelbreth–Holm–Swarm murine sarcoma basement membrane laminin (Sigma). The neural progenitor stage was maintained in reduced modified medium that consisted of DMEM:F12 with 15mM HEPES and sodium bicarbonate (Sigma), supplemented with 0.03% human serum albumin (GE Healthcare Life Sciences, Buckinghamshire, UK), 100 µg/mL apo-transferrin (Scipac Ltd, Kent, UK), 16.2 µg/mL putrescine (Sigma), 5 µg/mL human insulin (Sigma), 60 ng/mL progesterone (Sigma), 2 mM GlutaMAX™ Supplement (ThermoFisher Scientific), 40 ng/mL sodium selenite (Sigma), 10 ng/mL human fibroblast growth factor (FGF) (PeproTech, Rocky Hill, NJ, USA), 20 ng/mL human epidermal growth factor (EGF) (PeproTech) and 100 nM 4-OHT (Sigma). Cells were fed every 2-3 days and passaged using Accutase (Sigma) once 80–90% confluency was achieved. NPCs derived from hiPSC were obtained from keratinocytes of an apparently healthy individual and

grown in 6-well plates that had been previously coated with Geltrex (Life Technologies). The hiPSC lines were generated from primary keratinocytes as described elsewhere (Cocks et al., 2014), and their neuronal differentiation was stimulated by replacing E8 medium on confluent (>95%) hiPSCs with neuralisation medium (1:1 mixture of N2- and B27-containing medium, supplemented with 5 µg/ml insulin, 1 mM l-glutamine, 100 µm non-essential amino acids, and 100 µM 2-mercaptoethanol) supplemented with 1 µm Dorsomorphin (Sigma) and 10 µm SB431542 (Cambridge Bioscience, Cambridge, UK), as performed by Shi et al. (2012). At day 8, accompanied by the appearance of neuroepithelial cells, cells were passaged using Accutase (Sigma) and maintained in neuralisation medium only. Neuroepithelial cells were grown and passaged three times until neural rosettes (NPCs, ~day 18/19) were found. The NPCs were fed daily and passaged using Accutase (Sigma) once 80–90% confluency was achieved.

4.3.2. Cationic lipid-mediated vector transfection

For immunocytochemistry experiments, cells were grown on 13 mm coverslips coated on Geltrex (hiPSC-derived NPCs, n=1) or Laminin (CTX0E16 NPCs, n=1), which had been previously treated with 1M hydrochloric acid, 70% ethanol and pure ethanol in one hour washes each, followed by overnight sterilisation (180ºC) in a dry oven. Coverslips were placed in 24-well plates and approximately 10,000 cells were seeded per well the day before transfection. The next day, media was fully replaced by a final volume of 250 µL/well, and cells were transfected with plasmids for 24 hrs with the addition of 50 µL transfection solution/well previously incubated at room temperature for 45 min. This transfection solution consisted of 2% (v/v) Lipofectamine 2000 (ThermoFischer Scientific) and 1 µg DNA diluted in Opti-MEM Reduced Serum Media (ThermoFischer Scientific). In order to collect more protein for western blotting, HEK293T cells (n=3) were transfected in 6-well plates in a total volume of 2.5 mL per well, where 2 mL consisted of fresh medium plus 0.5 mL solution consisting of 1% (v/v) Lipofectamine 2000 and 2.5 µg DNA diluted in

Advanced DMEM/F-12. Plasmids used in this study included pEGFP-C3 (GeneBank ID U57607) and NT5C2-Myc-DDK (Origene ID RC200194). Water was used as a negative control in parallel with transfections in all experiments.

4.3.3. Protein extraction and quantification

Total protein was extracted from cell cultures by mechanical disruption of cells in RIPA Buffer (ThermoFisher Scientific) containing 1x Halt™ Protease Inhibitor Cocktail (ThermoFisher Scientific). Cell lysate was sonicated for 30 sec with a 50% pulse and incubated at 4ºC for 15 min. Samples were centrifuged at 14,000 ×g for 15 min to pellet cell debris, and the supernatant was transferred to a new tube and stored at -80ºC. Protein extracts were quantified using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific), where the reduction of copper by proteins in the solution in an alkaline medium is coupled to the colorimetric detection of cuprous cations (Cu^{1+}) by bicinchoninic acid (BCA). Absorbance at 595 nm was estimated using 96-well plates in a Beckman Coulter DTX 880 Multimode Detector, and protein concentrations were estimated using a standard curve of eight serial dilution points of bovine serum albumin solutions.

4.3.4. SDS-PAGE and Western Blotting

Protein samples were diluted to 1x Laemmli sample buffer (Bio-Rad, Hercules, California, USA) containing beta-mercaptoethanol. Samples were loaded in 15-well 4–20% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad). Electrophoresis was run at 150 V for approximately one hour. Proteins were wet-transferred to PVDF membranes for 90 V during 90 min. Membranes were blocked with Odyssey® Blocking Buffer (LI-COR Biosciences, Lincoln, NE, USA) for one hour and incubated overnight at 4ºC with primary antibodies diluted in blocking solution. Primary antibodies tested to probe NT5C2 were used at 1:200 dilutions and included the rabbit polyclonal ab96084 (Abcam, aa 108-332, Cambridge, UK) and the mouse

monoclonal M02-3C1 (H00022978-M02; Abnova, Taipei, Taiwan). The mouse monoclonal 9e10 Myc antibody (Evan et al., 1985) was used at 1:1,000 dilutions and was purchased from the NIH Developmental Studies Hybridoma Bank at The University of Iowa, USA. Following the overnight incubation of primary antibodies, membranes were washed three times for 10 min in TBS-T, and exposed to fluorescent secondary antibodies diluted 1:10,000 (Alexa Fluor 680 goat anti mouse, and Alexa Fluor 790 goat anti rabbit, ThermoFisher Scientific) for one hour. The membranes were washed three times again (10min each wash) and scanned using Odyssey® (LI-COR), where black and white images were acquired.

4.3.5. Immunocytochemistry and image acquisition

Fluorescently labelled cells were observed under confocal microscopy in order to investigate expression of endogenous NT5C2 and of overexpressed NT5C2-Myc/GFP. After 24 hrs transfections with plasmids or negative control, cells were double fixed using 10 min incubations with 4% formaldehyde in Phosphate-Buffered Saline (PBS) (ThermoFisher Scientific) containing 4% sucrose, and then cold absolute methanol on ice. Cells were permeabilised and blocked with 2% normal goat serum (NGS) in PBS-T (PBS containing 0.1% Triton-X). Primary antibodies tested included the previously mentioned NT5C2 antibodies (1:200), the Myc antibody (1:750) and the chicken ab13970 GFP antibody (1:1,000). These were incubated overnight in blocking solution at 4ºC and washed three times (3x10 min) with PBS-T on the next day. Cells were incubated with fluorescent secondary antibodies (Alexa Fluor goat anti-chicken 488, goat anti-mouse 568, goat anti-rabbit 633, ThermoFisher Scientific) in blocking solution for one hour. Cells were washed for three times and stained with DAPI (1:50,000) during 5 min. Coverslips were mounted using ProLong Gold Antifade Mountant reagent (ThermoFisher Scientific) and kept protected from light. Fluorescently labelled cells were imaged in a Leica SP5 confocal microscope using a 63x objective. Settings including laser power, pinhole and smart gain were kept constant within experiments. These were taken as z-stacks containing 9 or 10

plans (z-step = 0.5µm), and exported to ImageJ (Schneider et al., 2012) where background subtracted images and maximum intensity projections were generated (Srivastava et al., 2011).

4.3.6. Brain sections

Human *post-mortem* dorsolateral prefrontal cortex sections embedded in paraffin were obtained from two subjects free from psychiatric or neurological diagnosis at the time of death, from the MRC London Neurodegenerative Diseases Brain Bank at the Institute of Psychiatry, Psychology and Neuroscience, King's College London (UK). Samples included a 76-year-old female and an 80-year-old male. Ethical approval for this study was provided by The Joint South London and Maudsley and The Institute of Psychiatry NHS Research Ethics Committee (reference PNM/12/13-102).

4.3.7. Immunohistochemistry and image acquisition

Paraffin-embedded sections were deparaffinised and hydrated prior to immunostaining using 3,3'-diaminobenzidine (DAB) and Nissl stain. While DAB staining locates the protein of interest by means of immunoaffinity, Nissl counterstains the tissue by labelling RNA and DNA given their basophilic properties. Nissl stains strongly the granules of endoplasmic reticulum (ER) (called 'Nissl substance'), which allow the distinction between neurons (which will strongly stain due to the high amounts of ER in the soma, associated with intense protein synthesis), glial cells (mostly DNA is stained, as there is less protein synthesis, which leads to lower staining intensity), neuropil (synaptically dense areas that are too thin to visualise and contain very low number of cell bodies) and capillaries (very distinct endothelial cells surrounding an empty space) (Paul et al., 2008). Briefly, samples were deparaffinised and hydrated in a serial wash consisting of absolute xylene (5 min, twice), 99% industrial methylated spirit (IMS) (2 min, twice), 95% IMS

(2 min), and 70% IMS (2 min). As DAB staining is given by the oxidation of diaminobenzidine by peroxidases bound to the primary antibody and producing colour, endogenous peroxidases were blocked prior to the immunostaining. For this, sections were incubated in methanol solution containing 0.8% (v/v) hydrogen peroxide. Sections were then washed in distilled water and submitted to an antigen retrieval step. This step ensures epitopes are exposed and therefore able to bind to the primary antibodies. Sections were immersed in a citrate buffer solution (pH 6) and microwaved for 6 min on high, and for 8 min on low power. Sections were blocked with 10% NGS in TBS during 20 min, and NT5C2 antibody M02-3C1 (Abnova, H00022978-M02) was added 1:200 in a 1% NGS solution in TBS. A section was incubated with water instead of primary antibody for use as negative control of staining. Sections were incubated overnight at 4ºC, and on the next day washed twice in TBS, for 10 min. Secondary antibody (biotinylated rabbit anti-mouse immunoglobulin) was added 1:100 for 45 min, and the sections were washed again. Avidin/biotinylated horseradish peroxidase complex Vectastain® (Vector Laboratories, Peterborough, UK) was used for amplification and detection of the signal, according to the manufacturer's instruction. The sections were incubated for approximately 5 min with activated DAB solution (filtered and added 2 µL hydrogen peroxide per 5 mL of solution), and then washed in running tap water for 10 min. Slides were then sunk in a filtered aqueous solution of 0.5% cresyl violet containing 10% glacial acetic acid for approximately 15 min, at 55ºC. Sections were dehydrated in a serial wash consisting of, successively, 70% IMS (2 min), 95% IMS (2 min), 99% IMS (2 min, twice), absolute xylene (5 min, twice). The slides were mounted using DPX Mountant for histology (Sigma). Bright field images were taken using the 5x and 40x objectives, a Leica DMRB microscope, and the software AxioVision v4.8.2.

4.4. Results

4.4.1. Validation of NT5C2 antibodies

In order to investigate the role of NT5C2 in the nervous system, its distribution in the adult brain and in human cellular models is described in this chapter. For this, two commercially available antibodies for NT5C2, M02-3C1 and ab96084, were initially tested for specificity in western blots using extracts of HEK293T cells overexpressing an NT5C2 construct. This construct encodes the full length NT5C2 protein containing a C-terminal Myc-DDK tag, which should allow the co-localisation of immunostaining with multiple antibodies (for NT5C2 and Myc, for instance) (Figure 2). Single bands corresponding to the expected molecular weight of NT5C2 monomers (approximately 65 kDa) were detected using Myc and NT5C2 antibodies in the cell lysates of HEK293T cells overexpressing the construct. These results support the specificity of these antibodies for the exogenously expressed, tagged NT5C2 protein. None of the NT5C2 antibodies, however, were able to detect endogenous amounts of the protein in the cortical neural progenitor cells CTX0E16, the bone marrow cells SH-SY5Y, or the human embryonic kidney cells HEK293T. This suggested that the antibodies for NT5C2 would require time-consuming troubleshooting of electrophoresis and/or blotting methods, since the expression of *NT5C2* was detected by RT-PCR in RNA/cDNA derived from cultures from which protein extracts were obtained (data not shown).

Expression of NT5C2 at the protein level, therefore, was additionally investigated in two homologous cellular models of neural development, the NPCs CTX0E16, derived from the neuroepithelium of a human foetus, and the NPCs derived from hiPSCs obtained from a healthy subject, using immunocytochemical approaches. While the antibody M02-3C1 produced a staining pattern that diffused through the soma and early processes, the ab96084 antibody stained puncta-like structures in the soma and processes (Figure 3). Although both patterns reflected a ubiquitous presence of NT5C2 in the cytosol, it was unclear which staining profile corresponded to the true staining of NT5C2.

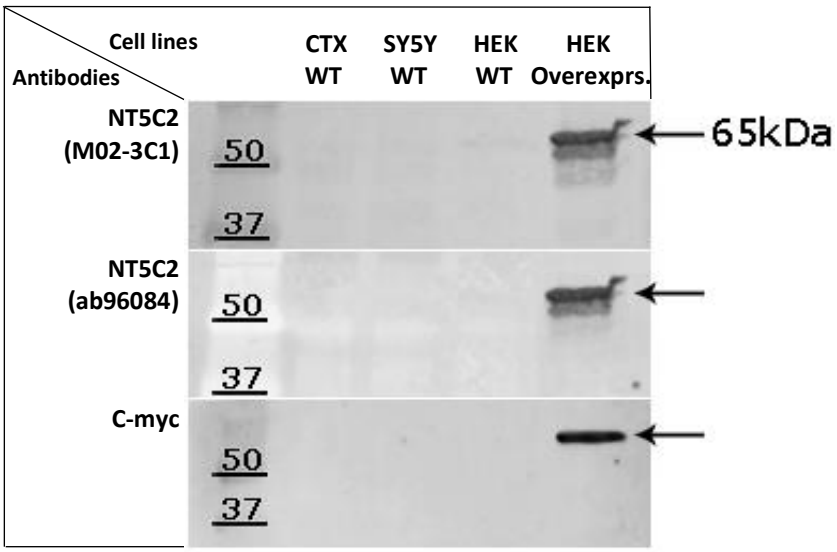


Figure 2. Immunoblotting for NT5C2 and Myc in protein lysates from multiple cell lines. These cellular lysates were obtained from non-transfected (WT) CTX0E16, SH-SY5Y and HEK293T, and transfected HEK293T, exogenously expressesing Myc-tagged NT5C2. No NT5C2 antibody was found to be sensitive for the detection of endogenous NT5C2 under these experimental conditions, but both NT5C2 antibodies were able to detect exogenous NT5C2 (Myc-tagged).

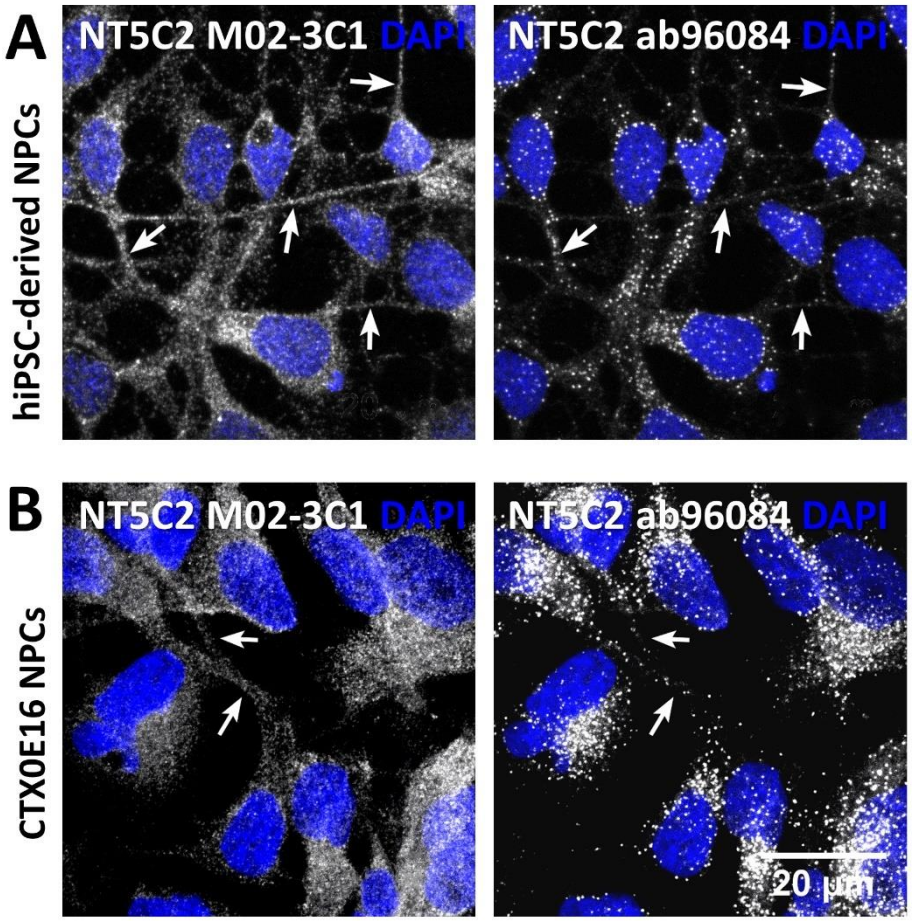


Figure 3. Immunostaining for NT5C2 in NPCs using two antibodies. **(A)** hiPSC-derived NPCs, **(B)** CTX0E16 NPCs. Immunofluorescence images were taken using a confocal microscope. The M02-3C1 (Abnova) antibody resulted in a diffuse distribution of NT5C2, whereas the ab96084 (aa 108-332) antibody stained puncta-like structures, suggesting they probe for different epitopes. Cellular processes are indicated by white arrows.

In order to address this uncertainty and to identify the most suitable antibody for NT5C2 immunostaining, hiPSC-derived NPCs were co-transfected with constructs encoding NT5C2-Myc and GFP. Examination of the immunostaining patterns obtained using the NT5C2 antibody ab96084 in parallel with antibodies for Myc and GFP (Figure 4A) revealed that ab96084 did not co-localise with exogenously expressed NT5C2-Myc, indicating its fluorescence signal to be unspecific. In parallel, similarly co-transfected cells were immunostained with NT5C2 antibodies M02-3C1 and ab96084, and a GFP antibody (Figure 4B). The resulting images obtained demonstrated that the M02-3C1 antibody co-localised with endogenous and exogenously expressed NT5C2. Immunostaining patterns observed for exogenous and endogenous NT5C2 using the M02-3C1 antibody were identical, with the fluorescence signal found diffused through the soma and cellular processes (Figures 3, 4B, 5B).

These findings were replicated using the CTX0E16 cells. Examination of the immunostaining pattern obtained using the NT5C2 antibody ab96084 in parallel with antibodies for Myc and GFP (Figure 5A) indicated that ab96084 was not suitable for immunocytochemistry staining. In parallel, similarly co-transfected cells immunostained with the NT5C2 antibodies M02-3C1 and ab96084, and a GFP antibody (Figure 5B), suggested that the M02-3C1 antibody successfully co-localised with endogenous and exogenously expressed NT5C2. The immunostaining patterns observed for endogenous and overexpressed NT5C2 using the M02-3C1 antibody in CTX0E16 cells were identical, with fluorescence diffused through the soma and cellular processes (Figure 5B). Collectively, these data suggest that the M02-3C1 antibody is adequate for NT5C2 labelling in immunocytochemistry and western blot experiments.

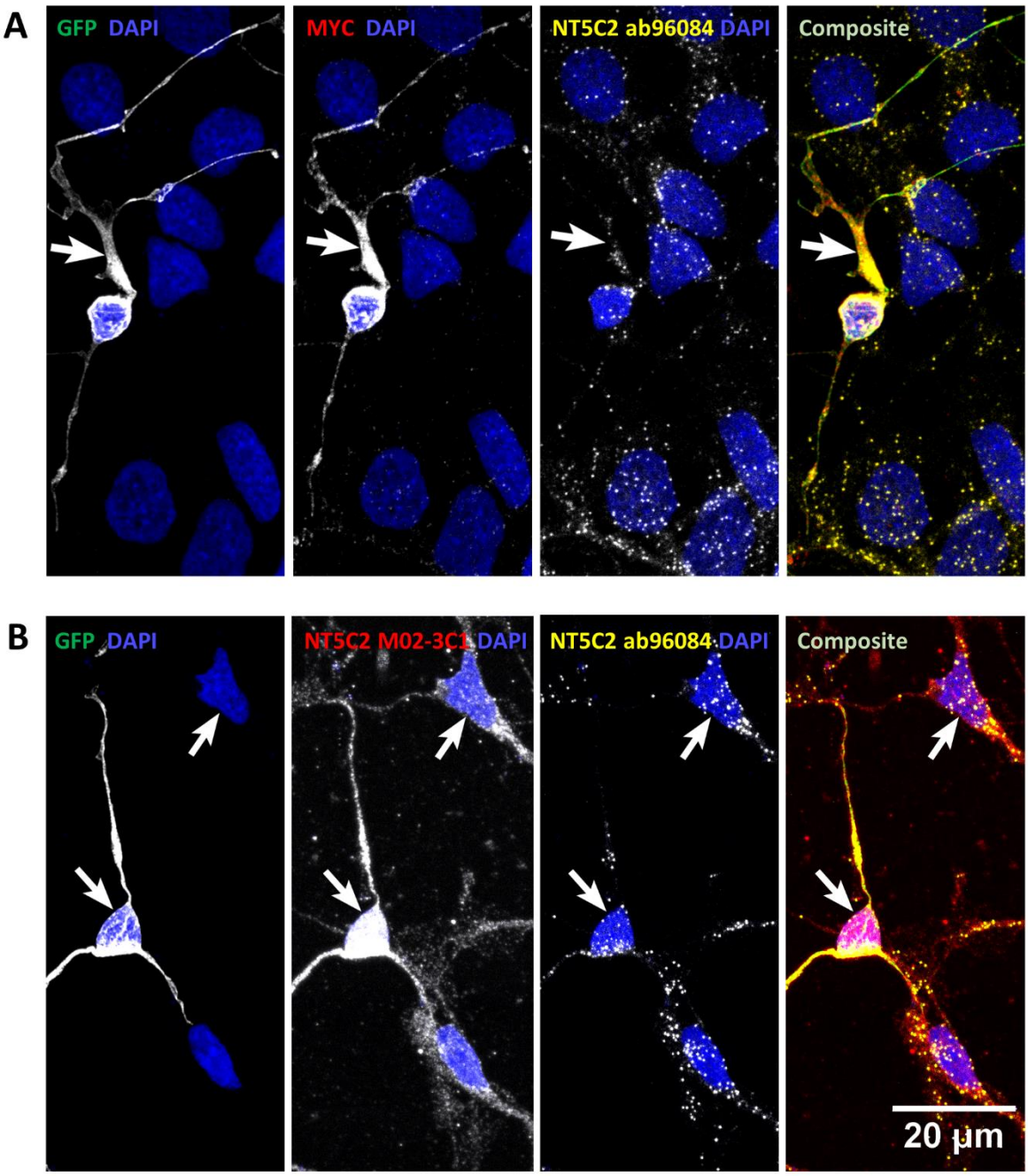


Figure 4. Validation of NT5C2 antibodies using immunocytochemistry in hiPSC-derived NPCs. Confocal microscopy images of NPCs co-transfected with constructs encoding GFP and NT5C2-Myc. **(A)** Co-localisation of staining for GFP with Myc (arrow), but not with ab96084 (NT5C2) in a co-transfected cell. In the composite, fluorescence signals associated with GFP (green), Myc (red) and NT5C2 ab96084 (yellow puncta) are overlaid. Individual channels are shown in black and white for better visualisation. **(B)** Co-localisation of staining for GFP with NT5C2 M02-3C1 antibody, but not with NT5C2 ab96084 (aa 108-332, central arrow) in a co-transfected cell. A non-transfected cell is indicated by the upper right arrow. In the composite, fluorescence signals associated with GFP (green), NT5C2 M02-3C1 (red) and NT5C2 ab96084 (yellow puncta) are overlaid. These data suggest that the NT5C2 antibody M02-3C1 is adequate for immunostaining.

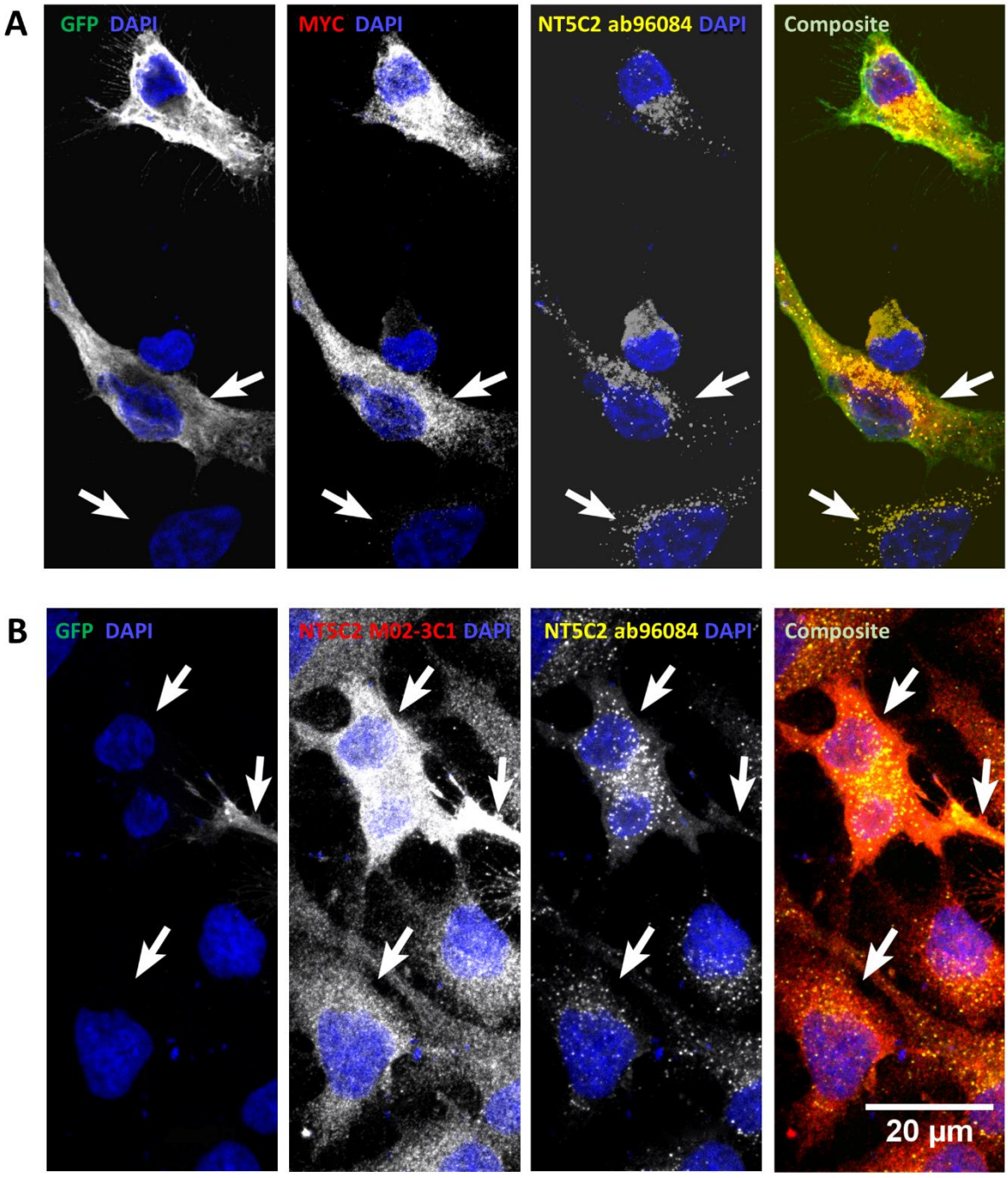


Figure 5. Validation of NT5C2 antibodies using immunocytochemistry in CTXOE16 NPCs. Confocal microscopy images of NPCs co-transfected with constructs encoding GFP and NT5C2-Myc. These results are consistent with findings using the hiPSC-derived NPCs. **(A)** Co-localisation of staining for GFP with Myc (mid arrow), but not with NT5C2 ab96084, in two co-transfected cells. In the composite, fluorescence signals associated with GFP (green), Myc (red) and NT5C2 ab96084 (yellow puncta) are overlaid. **(B)** Co-localisation of staining for GFP with NT5C2 M02-3C1, but not with NT5C2 ab96084 (aa 108-332, upper arrows), in co-transfected cells. Non-transfected cells are indicated by the lower arrow. In the composite, fluorescence signals associated with GFP (green), M02-3C1 (red) and ab96084 (yellow puncta) are overlaid. These data support the NT5C2 antibody M02-3C1 as the most adequate for immunostaining in this study.

4.4.2. Visualisation of NT5C2 in the adult DLPFC

The M02-3C1 antibody was used to interrogate the absolute expression of NT5C2 in the human DLPFC by means of immunohistochemistry using DAB and Nissl staining. Expression of NT5C2 was found to co-localise with neurons, glial cells and neuropil (synaptically dense region containing a relatively low number of cell bodies), consistent with the distribution of a ubiquitous enzyme. Moreover, staining of the neuropil is consistent with immunostaining of cellular processes of the NPCs. Some glial cells, however, did not co-localise with NT5C2 staining (Figure 6). Taken together, these data demonstrate that within the human DLPFC, pyramidal neurons express NT5C2, while a subset of glia cells express this protein. These data suggest that the NT5C2 protein could be differentially distributed through the cortex. Quantitative studies to identify whether pyramidal neurons of specific cortical layers or specific classes of glia cells express NT5C2 are warranted to clarify these findings.

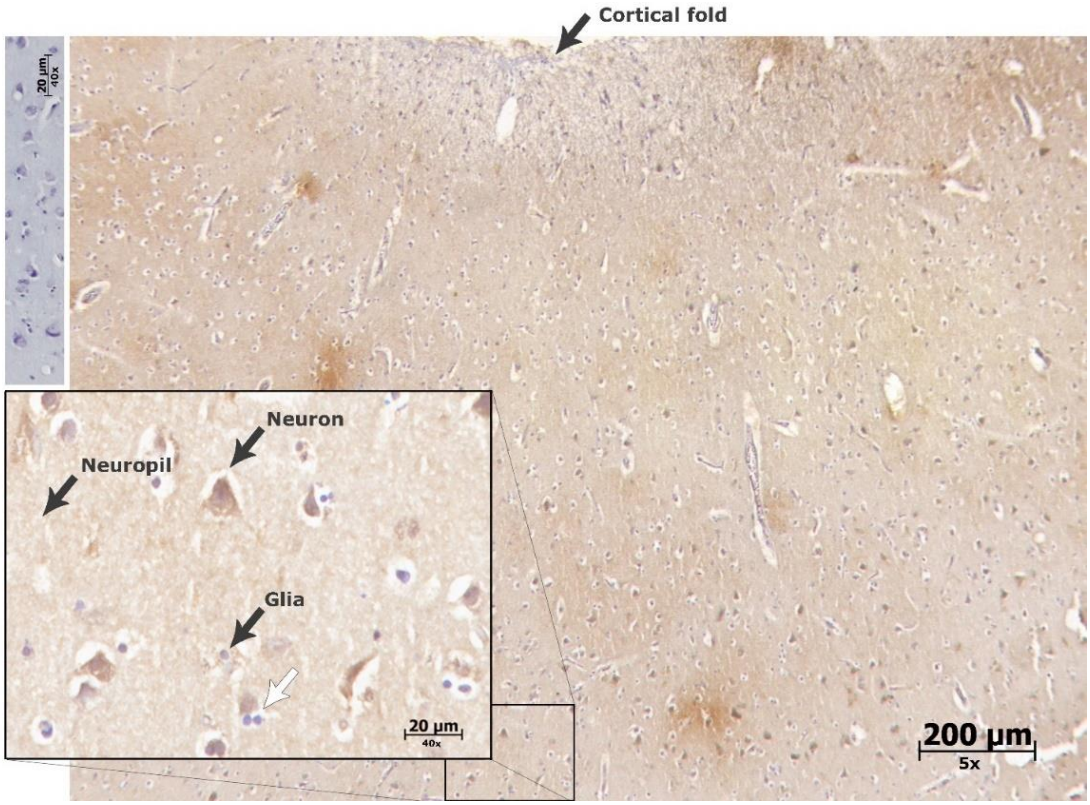


Figure 6. Immunohistochemistry of the human DLPFC probed for NT5C2. Immunostaining was performed using 3,3′-diaminobenzidine (DAB) and the NT5C2 M02-3C1 antibody. Sections were counterstained for Nissl substance using cresyl violet. Representation of the negative control (no primary antibody) is shown on the left top panel. Magnification is displayed under scale bars. Orientation of the image is given by the upper arrow showing a cortical fold. Lower black arrows point to examples of neurons, neuropil and glial cells stained for NT5C2, and a white arrow point to an example of a glial cell where NT5C2 staining was absent.

4.5. Discussion

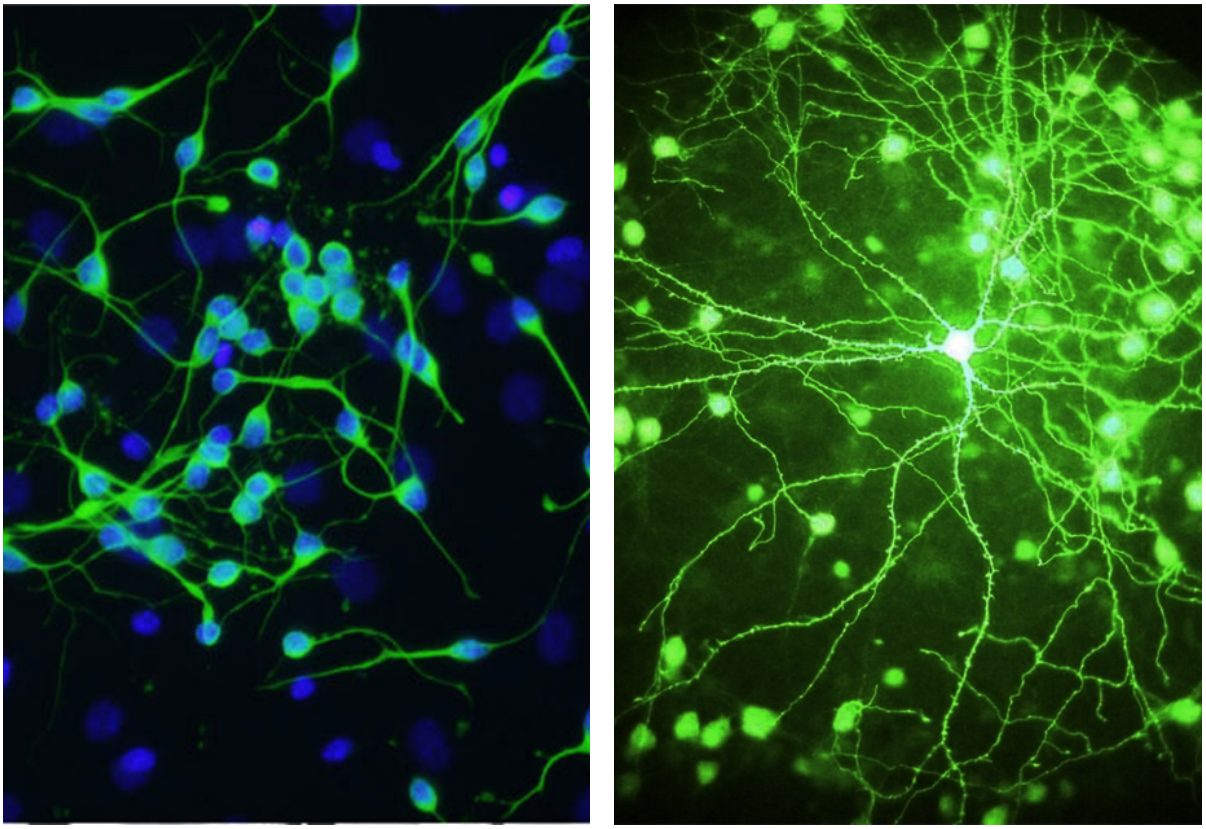
In Chapter 2, *cis*-regulatory effects operating on *NT5C2* in different brain tissues were attributed to genotype at the top risk variants associated with schizophrenia on chromosome 10q24. The majority of *cis*-regulatory effects for *NT5C2* in the DLPFC could be accounted for by genotype at these risk variants. The DLPFC is implicated in pathophysiological abnormalities observed in schizophrenia (Callicott et al., 2000). In Chapter 3, evidence for the existence of less expressed transcripts of *NT5C2* was found. The full length transcripts, however, seemed to be more highly expressed and thus potentially contributing more to the cis-regulatory effects detected in Chapter 2. As not much is known about the presence of *NT5C2* in the human brain, its distribution in the adult DLPFC and in NPCs derived from cortical origin was investigated in this chapter. Briefly, the cytosolic 5'-nucleotidase II (NT5C2) was confirmed as ubiquitously expressed in the investigated tissues.

The biggest issue with immunoaffinity-based experiments is the lack of antibody specificity, which has often led to wrong biological interpretations (Baker, 2015; Gilda et al., 2015). Two commercially available antibodies for NT5C2 (the rabbit polyclonal Abcam ab96084 aa 108-332 and the mouse monoclonal Abnova M02-3C1) were tested for immunostaining of endogenous and exogenous NT5C2 in different cellular models, using western blotting and immunocytochemistry. Both NT5C2 antibodies were shown to specifically detect single bands in western blotting using cell lysates of HEK293T cells transfected with this construct (Figure 2). Endogenous NT5C2 production in SH-SY5Y, CTX0E16 or HEK293T cells, however, was not detected under the experimental conditions in this chapter. It is possible that this event reflects low protein concentration or low sensitivity of the antibodies, which would require time-consuming optimisation of electrophoresis and blotting methods, and possibly acquisition of new antibodies.

Instead, neural progenitor cells derived from hiPSCs and from the human foetal brain (the CTX0E16 cells) were used to co-transfect plasmids encoding NT5C2-Myc and GFP, so that these cells could be immunolabelled using antibodies for NT5C2,

GFP and Myc. These experiments revealed that the antibody M02-3C1 was the most adequate for NT5C2 staining, while it informed of the spatial distribution of endogenous and exogenous NT5C2 protein in the NPCs. NT5C2 localisation in these cell lines was observed to be dispersed throughout the soma and cellular processes, which suggests this is a ubiquitous cytosolic enzyme (Figure 3) (Weirich et al., 2008). Moreover, the M02-3C1 antibody was used to investigate the profile of expression of NT5C2 in the DLPFC. Neurons, glial cells and the neuropil were found to be positive for NT5C2 immunostaining, although some glial cells that do not express NT5C2 were also identified. Quantitative immunohistochemistry data for NT5C2 in the cerebral cortex performed by Uhlen et al. (2015) suggested that >75% of neuronal cells and the neuropil successfully co-localised with NT5C2 staining, whereas a slightly smaller proportion of glial cells, between 25-75%, was found to express NT5C2. This could indicate that the risk mechanisms driving association with schizophrenia could be arising from cells which more abundantly express NT5C2.

Limitations of this study include the small sample size analysed, and the lack of a cell line or tissue that is known to constitutively not express NT5C2 (a negative control for staining). Investigating specific cell types involved in the cis-regulatory risk mechanism could be performed using single-cell RNA-seq or tissue microdissection or, alternatively, using quantitative immunostaining or fluorescence *in situ* hybridisation (FISH). Nevertheless, the results obtained support the use of the NT5C2 antibody M02-3C1 for *in vitro* or *post-mortem* immunolabelling, under the experimental conditions described in this chapter. In summary, this chapter found supporting evidence for the ubiquitous expression of NT5C2 in the central nervous system (Bianchi and Spychala, 2003; Itoh, 2013), in which neurons, neuropil and glial cells were found to co-localise with NT5C2 staining. Evidence in the literature that suggests that NT5C2 is more abundantly expressed in neurons rather than glial cells is partly supported by the present data. Further investigation is warranted, as this may have implications for the putative role of NT5C2 in risk for schizophrenia.



Rodent neural stem cells (left picture, by Maria Morell, University of Michigan) and a mature primary neuron (right picture, by Pooja Raval, King's College London). The stem cells were stained for morphological (green) and nuclear (blue) markers, and the adult neuron was transfected and stained for green fluorescent protein (GFP).

Chapter 5

Global gene expression profiling of a neural progenitor cell line following *NT5C2* knockdown

5.1. Abstract

Genome-wide association signals for schizophrenia on chromosome 10q24 were found to be associated with *cis*-regulatory effects on multiple genes at this locus, in Chapter 2. The risk alleles of rs11191419 and chr10_104957618_I were associated with decreased expression of *NT5C2* in three adult brain areas involved in schizophrenia pathophysiology and in the foetal brain, implicating *NT5C2* as a putative schizophrenia risk gene. Chapter 3 provided evidence suggesting that risk is more likely mediated via expression changes to full length *NT5C2*, while Chapter 4 found endogenous expression of this protein associated with all major neural cell types in the adult DLPFC. In order to study the effects on global gene expression elicited by this risk mechanism, the downregulation of *NT5C2* was reproduced *in vitro* using small interfering RNAs (siRNAs) in human neural progenitor cells (NPCs). Two independent siRNAs were found to effectively reduce *NT5C2* at the transcript and protein levels, as observed by qPCR and immunocytochemistry. The knockdown was also confirmed by western blot via measurement of AMPK-alpha Thr172 phosphorylation. The expression of 65 genes of known function was found to be significantly altered in association with the knockdown by the siRNAs. This gene list was enriched for GO and KEGG terms implicated in the regulation of cardiac muscle contraction and the actin cytoskeleton, as well as cancer (FDR<0.05). Enrichment for genes involved in the mitogen-activated protein kinases (MAPK) pathway was also observed at a more relaxed FDR threshold (FDR<0.1). Mean increased activation of MAPK signalling as function of phosphorylation of ERK1/2 was observed by western blotting, although this event was not statistically significant. The antipsychotics ziprasidone and fluphenazine were estimated by connectivity mapping to reverse the gene expression changes elicited by the knockdown in the NPCs. These data provide evidence of downstream mechanisms associated with the knockdown of *NT5C2,* potentially mediating risk to schizophrenia.

Note: The microarray presented in this chapter was performed by Mr Sanghyuck Lee, from the the BRC IoPPN Genomics & Biomarker Core Facility at the MRC Social, Genetic & Developmental Psychiatry Centre (SGDP) at King's College London. The microarray and pathway analyses were assisted by Dr Timothy Powell and Dr Gerome Breen from the SGDP. The Western blots to quantify ERK1/2 phosphorylation in cultures submitted to *NT5C2* silencing were performed by Mr Iain Watson (Srivastava's Lab).

5.2. Introduction

The largest genome-wide association study of schizophrenia published to date identified 108 genomic loci associated with the disorder (Schizophrenia Working Group of the PGC, 2014), from which the third most significant locus is on chromosome 10q24. In Chapter 2, genotype at the top risk variants rs11191419 and chr10_104957618_I was associated with altered *cis*-regulation of multiple genes at the locus. The risk alleles, however, acted concordantly to decrease expression of *NT5C2* in multiple brain tissues. In the DLPFC, for example, a brain region implicated in schizophrenia pathophysiology (Galindo et al., 2016), risk alleles were associated with a mean 16% reduction in the expression of *NT5C2* (Chapter 2, Figure 8) (Duarte et al., 2016). Although novel transcripts of this gene were detected in the human brain, the full length variants were more highly abundant (Chapter 3), and thus more likely to contribute to the cis-regulatory mechanism observed. Immunostaining of the adult DLPFC (Chapter 4) found all neural cell types to co-localise with the cytosolic 5'-nucleotidase II (*NT5C2*).

The cytosolic 5'-nucleotidase II (NT5C2) is a soluble enzyme which cleaves inorganic phosphate from purine nucleotides such as inosine, adenosine and guanosine monophosphate molecules (IMP, AMP, GMP) (Itoh, 2013). It also catalyses the transfer of phosphate groups between nucleotides and nucleosides (Meyer et al., 2013). As key regulator of intracellular nucleotide and nucleoside pools, *NT5C2* might lend support to the purinergic hypothesis of schizophrenia (Hirota and Kishi, 2013; Lara and Souza, 2000; Yao et al., 2010). This hypothesis suggests that purine levels can affect glycine abundance in the synaptic cleft, which in turn modulates the effect of glutamate, subsequently affecting GABAergic and dopaminergic neurons, and thus changes to NT5C2 levels may potentially explain positive and negative symptoms of schizophrenia (Boison et al., 2012).

Although the knockdown of *NT5C2* has been associated with cell death (Careddu et al., 2008) and changes to cellular proliferation in astrocytoma cells (Cividini et al., 2015b), and energetic mobilisation in muscle cells (Kulkarni et al., 2011), its effects on the global gene expression profile of NPCs remain unknown. Moreover, the effect of *NT5C2* knockdown on the global gene expression profile needs to be investigated in the context of schizophrenia, therefore using a relevant cellular model. The endogenous expression of *NT5C2* was reduced in a neural progenitor cell line (CTX0E16) derived from the cortical neuroepithelium of a first trimester foetus (Anderson et al., 2015). With the use of small interfering RNAs (siRNAs), the transient knockdown of *NT5C2* was achieved and the effects on the global gene expression profile were assessed using Illumina BeadChip microarrays. Gene ontology analysis and connectivity mapping methods were applied to the gene list for which significant expression changes were shared between two distinct, non-overlapping siRNA sequences targeting *NT5C2*. The objective of this chapter is therefore to identify downstream cellular effects associated with reduction of *NT5C2* expression, which is associated with risk for schizophrenia.

5.3. Methods

5.3.1. Cell culture

The neural stem-cell line CTX0E16 was used to investigate global gene expression changes caused by the knockdown of the schizophrenia susceptibility gene *NT5C2*. This cell line was kindly provided by ReNeuron (www.reneuron.com) as part of a Material Transfer Agreement. The cell line was derived from a 12-week old foetal brain, as detailed elsewhere (Anderson et al., 2015; Pollock et al., 2006). Briefly, this karyotypically normal cell line was obtained from the cortical neuroepithelium of a human foetus, and conditionally immortalised by genomic incorporation of the c-mycER^{TAM} transgene. This construct acts as molecular switch to stimulate proliferation in the presence of 4-hydroxi-tamoxifen (4-OHT). Cells were maintained at 37ºC with 5% CO_2, and routinely grown in Nunclon™ Delta Surface T75 tissue culture flasks (ThermoFisher Scientific, Waltham, MA, USA) coated with 1 µg/cm^2

Engelbreth–Holm–Swarm murine sarcoma basement membrane laminin. Media was changed every 2-3 days, cells were passaged when 80-90% confluency was achieved, and cultures were maintained for between 13 and 20 passages. Proliferation was induced by maintaining cells in reduced modified medium (DMEM:F12 with 15 mM HEPES and sodium bicarbonate) supplemented with 0.03% human serum albumin (GE Healthcare Life Sciences, Buckinghamshire, UK), 100 µg/mL apo-transferrin (Scipac Ltd, Kent, UK), 16.2 µg/mL putrescine, 5 µg/mL human insulin, 60 ng/mL progesterone, 2 mM L-glutamine, 40 ng/mL sodium selenite, 10 ng/mL human fibroblast growth factor (FGF) (PeproTech, Rocky Hill, NJ, USA), 20 ng/mL human epidermal growth factor (EGF) (PeproTech) and 100 nM 4-OHT. For immunocytochemistry, cells were grown on laminin-coated 13 mm coverslips, which had been previously treated with 1 M hydrochloric acid, 70% ethanol and pure ethanol for one hour each, followed by overnight sterilisation (180ºC) in a dry oven. All reagents were purchased from Sigma-Aldrich (Gillingham, Dorset, UK) unless stated otherwise.

5.3.2. Small interfering RNA (siRNA) in cultures

Double-stranded RNA molecules (dsRNA) can be transfected into cells and converted into siRNA molecules owing to the endogenously expressed enzyme Dicer. The siRNA molecule associates with the RNA Induced Silencing Complex (RISC) and Argonaute2 to cleave and/or inhibit translation of specific messenger RNAs to which they pair with by complementarity (Whitehead et al., 2009). In this study, the Trilencer 27-mer *NT5C2* siRNA kit (SR307908, Origene, Rockville, MD, USA) was used to transiently knockdown full length *NT5C2*, RefSeq transcripts NM_012229 and NM_001134373. dsRNA Resuspension Buffer (provided by the manufacturer) was used to solubilise the oligoribonucleotides, and the solutions were heated at 94ºC for 2 min after resuspension. The kit contained three non-overlapping siRNA sequences, which included: siRNAs "A" (UGAGAAGUAUGUAGUCAAAGAUGGA, sense sequence) and "B" (ACAACUGUAAUAGCUAUUGGUCUTC, sense sequence), equally targeting the utter 5' exon of both known *NT5C2* transcripts, and siRNA "C"

(CAAAGACACUGACUACAAGCGGCAC, sense sequence), targeting exon 10 from NM_012229 and exon 9 from NM_001134373, according to the human genome build hg19. The new transcripts of NT5C2 detected in Chapter 3 are expected to be similarly silenced by these siRNA sequences. The corresponding dsRNAs were transfected in parallel with (1) a negative control (scramble sequence) (SR30004, sense: CGUUAAUCGCGUAUAAUACGCGUAT); and (2) a fluorescent reagent, the positive control BLOCK-iT™ Fluorescent Oligo (ThermoFisher Scientific) to provide estimates of transfection efficiencies into the CTX0E16 cells by using fluorescence microscopy. A million cells were plated into T75 flasks 24 hours before transfection (or 10,000 cells for 13 mm coverslips placed in 24-well plates, for the immunocytochemistry assays). Four biological replicates (n=4) were transfected with 10 nM siRNAs in parallel with the controls. On the day of transfection, media was changed to reduced modified medium supplemented with all items with the exception of 4-OHT, so that proliferation was not artificially stimulated through c-Myc overexpression. Cells transfected with the positive control were manually counted in an Olympus IX70 microscope (10x objective) using bright field visualisation, a fluorescence source and a 568 nm filter to ascertain the level of transfection. Global gene expression was assessed only in the two siRNA conditions for which siRNAs yielded the best silencing in the transcript and protein level.

5.3.3. RNA and protein extractions

In order to extract total RNA and protein from each flask, cells were detached using Accutase, and the cell suspension was equally split into two tubes for harvesting. For total RNA extraction, one of the pellets was mechanically dissolved in 1 mL Tri-Reagent (ThermoFisher Scientific), as described in the manufacturer's protocol. Briefly, the resulting homogenate was mixed with 0.1 mL 1-bromo-3-phenolpropane, and incubated at room temperature for 15 min. The homogenate was then centrifuged at 13,000 xg for 10 min at 4ºC. The supernatant was transferred to a new tube containing 0.5 mL isopropanol, and was incubated at room temperature for 10 min. After harvesting, the supernatant was discarded and

the pellet was washed three times with molecular grade 75% ethanol. The supernatant was discarded and the sample was let to air dry for 15 min. The pellet was resuspended in nuclease-free H_2O, and absorbances were measured in a spectrophotometer ND-1000 (NanoDrop). For protein extraction, the pellet was manually disrupted in RIPA Buffer (ThermoFisher Scientific) containing 2x Halt Protease and Phosphatase Inhibitor Cocktail (ThermoFisher Scientific). The cell lysates were sonicated at 4ºC for three 30 sec pulses at 50% power, followed by 30 sec rest steps between bursts. The lysate was incubated at 4ºC for 15 min and harvested at 13,000 xg for 15 min at 4ºC to collect for cell debris. The supernatant was collected and stored at -80ºC.

5.3.4. DNAse treatment

Total RNA samples were treated with TURBO DNA-free kit™ (ThermoFisher Scientific) prior to use, according to the protocol provided by the manufacturer. Briefly, 10 µL RNA (approximately 4 µg) was incubated with 2 units of Turbo DNAse in 1x Turbo DNAse buffer at 37ºC for 30 min. Reaction inactivation was performed by incubation with DNase Inactivation Reagent during 5 min. The resulting solution was centrifuged and the supernatant (DNA-free RNA) collected. This RNA did not yield a PCR product in the absence of a reverse transcription step.

5.3.5. cDNA synthesis for quantitative PCR

Reverse transcription was performed with 1.5 µg of DNA-free RNA using SuperScript III® (SSIII) Reverse Transcriptase (ThermoFisher Scientific). Briefly, the reaction was performed in two steps that led to a 20 µL final reaction containing 1.5 µg RNA, 5 µM random decamers, 500 µM dNTPs, 5 mM DTT, 40 units RNaseOUT™ Recombinant Ribonuclease Inhibitor (ThermoFisher Scientific) and 200 units of SSIII enzyme in 1x First-Strand Buffer. An initial mixture containing RNA, random decamers and dNTPs was heated to 65ºC for 5 min and cooled on ice for at least 1 min, when the remaining components were added to the reaction. The reaction was incubated on a GS4 Thermocycler (G-Storm, Somerton, UK) with a heated lid for 5 min at 25ºC, 60 min at 50ºC, 30 min at 55ºC and 15 min at 70ºC. The resulting cDNA was diluted 1:7 for the RT-qPCR reactions and stored at -20ºC.

5.3.6. Quantitative PCR (qPCR)

Quantitative PCR was carried in 12.5 µl reactions consisting of 5 µL 1:7 diluted cDNA, 1× HOT FIREPol EvaGreen qPCR Mix Plus (Solis BioDyne, Tartu, Estonia) and 200 nM primers using a real-time thermocycler MJ Research Chromo 4 (Bio-Rad, Hercules, CA, USA) linked to an MJ Opticon Monitor analytic software (Bio-Rad). The reaction programme consisted of a 15-min incubation at 95ºC and 40 cycles of 30 sec at 95ºC and 30 sec at 60ºC. Duplicate qPCR reactions were performed, and standard deviation between cycle thresholds (CTs) for technical replicates was <0.5. The reactions yielded single melting curves when submitted to a 60ºC-95ºC temperature gradient. The oligonucleotides shown on Table 1 were designed using Primer3 (Untergasser et al., 2012), and were purchased from Integrated DNA Technologies (London, UK). All primer sets had PCR efficiencies calculated as determined by a standard curve of four pooled cDNA dilution points. The primers used for validating the microarray findings were designed to amplify close to the exon complementary to the microarray probes. Gene expression was estimated using the Pfaffl (2001) method, which takes into account CTs and efficiencies calculated for target and reference genes. Gene expression values were first normalised to *B2M* and *RPL13A* levels (see below), then averaged and compared to expression values at the scramble condition (negative control) using two-tailed t-tests on GraphPad Prism 6 (GraphPad Software, La Jolla, CA, US).

5.3.7. Housekeepers screening

A screening for the most stably expressed housekeeping genes (HKs) upon knockdown of *NT5C2* was performed in cultures submitted to all siRNA treatments. The tested HKs included *B2M* (beta-2-microglobulin), *RPL13A* (ribosomal protein L13a), *ACTB* (beta-actin), *SDHA* (succinate dehydrogenase complex, subunit A), *UBC* (ubiquitin C), *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) and *RPL30* (ribosomal protein L30) (Table 1). The most stable housekeepers were selected using NormFinder (Andersen et al., 2004), which takes into consideration the expression stability between conditions, and PCR efficiencies inherent to each primer set. The most stable expression was observed for *B2M* and *RPL13A*, as shown in Figure 1

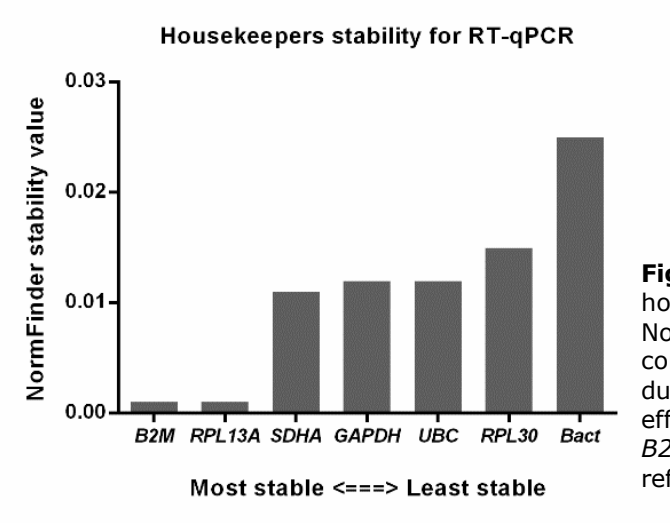


Figure 1. Expression stability of tested housekeeping genes, according to NormFinder. This analysis takes into consideration expression stability during the knockdown and the PCR efficiencies inherent to each primer set. *B2M* and *RPL13A* were chosen as reference genes for RT-qPCR assays.

Table 1. Oligonucleotide sequences used for qPCR assays in this study.

Target	Forward 5'-3'	Reverse 5'-3'	Usage
NT5C2	CTCCCAACCTCTTCCCACTG	GGACCTCGTTTGTTCCTGTG	NT5C2 expression
ATG4B	TACTCTGACCTACGACACTCTCC	TGCTGTATTTTCTACCCAGTATCCA	Validation microarray
PSMC4	GTCCTATCCTGCCCTTCCAG	GTGTGGCCTGGGATGATCT	Validation microarray
HNRNPA1	CGGAGTCTACCAATGCCGAA	CAGAAAGGAGCAAGCTGACG	Validation microarray
TBCA	AAATATGCGGCTTCCAACCTG	AATGAGAGCTGAAGACGGTGA	Validation microarray
B2M	TATCCAGCGTACTCCAAAGA	GACAAGTCTGAATGCTCCAC	Housekeeper screen (stable)
RPL13A	CCTGGAGGAGAAGAGGAAAGAGA	TTGAGGACCTCTGTGTATTTGTC	Housekeeper screen (stable)
ACTB	TGTGATGGTGGGAATGGGTCAG	TTTGATGTCACGCACGATTTCC	Housekeeper screen (unstable)
SDHA	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCATG	Housekeeper screen (unstable)
UBC	ATTTGGGTCGCGGTTCTTG	TGCCTTGACATTCTCGATGGT	Housekeeper screen (unstable)
GAPDH	GAAATCCCATCACCATCTTCCAGG	GAGCCCCAGCCTTCTCCATG	Housekeeper screen (unstable)
RPL30	ACAGCATGCGGAAAATACTAC	AAAGGAAAATTTTGCAGGTTT	Housekeeper screen (unstable)

5.3.8. Immunocytochemistry

The validation of the NT5C2 antibody M02-3C1 was presented in Chapter 4 (the monoclonal antibody M02, clone 3C1, H00022978-M02, Abnova, Taipei, Taiwan). In this chapter, this antibody was used to quantify the knockdown of *NT5C2* at the protein level using immunocytochemistry experiments in independent cultures.

Cultures were incubated with 10 nM siRNAs for 72 hours, after which cells were double-fixed using 4% formaldehyde in phosphate-buffered saline (PBS) containing 4% sucrose for 10 min, at room temperature. Cells were then incubated with cold methanol for 10 min on ice. Fixed cells were permeabilised and blocked with 2% normal goat serum (NGS) in PBS-T (PBS containing 0.1% Triton-X). The antibody for NT5C2 (mouse, M02-3C1, 1:200) and for the morphological marker beta-3 tubulin (Abcam ab41489, aa 108-332, chicken, 1:500) were incubated overnight in blocking solution at 4ºC. On the next day, cells were washed for 10 min three times in PBS-T and incubated with fluorescent secondary antibodies (Alexa Fluor goat anti-mouse 488 and goat anti-chicken 568, ThermoFisher Scientific) in blocking solution for one hour. Cells were washed again and stained with DAPI (1:50,000) for 5 min. Coverslips were mounted using ProLong Gold Antifade Mountant reagent (ThermoFisher Scientific), and stored at 4ºC protected from light.

5.3.9. Confocal imaging acquisition and analysis

Fluorescently stained CTX0E16 cell were imaged in a Leica SP5 confocal microscope using the 40x objective. Settings like laser power, pinhole and smart gain were adjusted during the imaging of the control, and were kept constant during the image acquisition. Images were taken as z-stacks containing 10 plans (z-step = 0.5 µm) and exported to ImageJ (Schneider et al., 2012), where background subtracted images and maximum intensity projections were generated (Srivastava et al., 2011). The beta-3 tubulin channel was thresholded to delineate the cell soma and mark regions of interest (ROIs). All images were converted to greyscale, and corrected total cell fluorescence (CTCF) for NT5C2 was calculated as: *CTCF = integrated density - (area x mean fluorescence of three background reads).* Each condition was imaged in four fields of view (FOV) and CTCF values were obtained for 20 cells per FOV, in four biological replicates (approximately 320 cells/condition). Resulting CTCF values were compared in GraphPad Prism 6 using a one-way ANOVA.

5.3.10. Protein quantification and Western blotting

Protein extracts were quantified using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific), as described in the previous chapter. Briefly, this assay couples the reduction of copper by proteins in an alkaline medium, with colorimetric detection of cuprous cations (Cu_{1+}) by bicinchoninic acid (BCA). Absorbance at 595 nm was estimated using 96-well plates in a Beckman Coulter DTX 880 Multimode Detector. Protein concentrations were estimated using a standard curve of eight serial dilution points of bovine serum albumin solutions. Protein samples were diluted to 1x Laemmli sample buffer (Bio-Rad, Hercules, California, USA) containing beta-mercaptoethanol. Approximately 5 µg total protein was loaded in 15-well 4–20% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad). Electrophoresis was adjusted to 150 V for approximately one hour. Proteins were wet-transferred to nitrocellulose membranes for 100 V during one hour. The membranes were blocked with Odyssey® Blocking Buffer (LI-COR, Lincoln, NE, USA) or milk 5% in TBS-T for one hour, and incubated overnight at 4ºC with primary antibodies in blocking solution. Primary antibodies used include the MAPK antibody (ERK1/2 antibody #9102, 1:1,000 in Odyssey buffer, Cell Signalling Technology, Danvers, MA, US), phospho-MAPK (Phospho-ERK1/2 Thr202/Tyr204 antibody #9101, 1:1,000 in Odyssey buffer, Cell Signalling Technology), AMPK-alpha (AMPK-alpha D6 antibody sc-74461, 1:200 in milk blocking buffer, Santa Cruz Biotechnology, Dallas, TX, USA) and phospho-AMPK-alpha (p-AMPK alfa1/2 Thr 172, sc-33524, 1:200 in milk blocking buffer, Santa Cruz Biotechnology). On the next day, membranes were washed three times for 10 min in TBS-T (20 mM Tris, 150 mM NaCl containing 0.1% Tween-20), and exposed to fluorescent secondary antibodies diluted 1:10,000 for one hour (Alexa Fluor 680 goat-anti-mouse, and Alexa Fluor 790 goat-anti-rabbit, ThermoFisher Scientific). The membranes were scanned in an Odyssey (LI-COR) machine after three 10 min washes. Regions of interest corresponding to bands of total-/phospho-MAPK were drawn and quantified using the LI-COR® Odyssey® Software (LI-COR). Integrated intensity values for phosphorylated MAPK and AMPK members were normalised to

total kinase amounts, normalised to phosphorylation at the control condition, and analysed in a non-parametric analysis of variance using GraphPad Prism.

5.3.11. Microarray

The cultures for which a significant knockdown of *NT5C2* was achieved and the control cultures treated with a scramble siRNA (siRNA A, n=3; siRNA B, n=4; controls, n=4) were submitted for microarray analysis at the IoPPN Genomics & Biomarker Core Facility, Social, Genetic & Developmental Psychiatry Centre, King's College London, UK, in collaboration with Dr Gerome Breen's group. Prior to microarray library preparation, RNA integrity was checked in all 14 samples using the Agilent RNA 6000 Nano Kit on the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) (RIN=10 for all samples) (Figure 2). Global gene-expression profiling was performed using the Illumina Human HT12 v4 BeadChip array (Illumina, Cambridge, Cambridgeshire, UK), according to the manufacturer's instructions.

Raw microarray probe fluorescence intensities were extracted from idat files using the GenomeStudio software. Spline normalisation and log2 transformation were applied to the data using the lumi Bioconductor package (Du et al., 2008) in RStudio. Probe intensities were analysed in a linear regression model to measure the effect of the siRNAs while correcting for variability between biological replicates and microarray batches. To limit spurious results, probes that were not detected in all samples with a detection threshold $P < 0.05$ were disregarded from the analysis. Differentially expressed genes ($P < 0.05$, uncorrected) showing the same direction of effect upon knockdown with two independent siRNA sequences targeting *NT5C2* were submitted to analysis of gene ontology and connectivity mapping. A hypergeometric probability was calculated to inform the probability of obtaining the resulting overlapping gene set from the differentially expressed gene lists in each siRNA condition, considering all genes expressed in the NPCs. Use of the overlapping gene set, as performed by Hill et al. (2012), is likely to remove the majority of off-target effects associated with individual siRNA sequences.

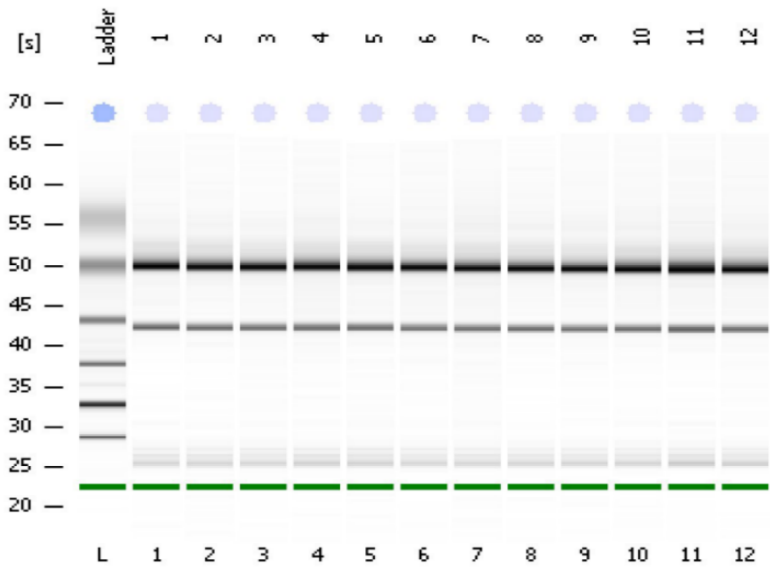


Figure 2. Quality control for total RNA using BioAnalyzer. Representative image. Analysis was performed on DNA-free RNA samples obtained from CTX0E16 cultures submitted to siRNA treatments A (n=3), B (n=4) and their respective controls, prior to microarray analysis.

5.3.12. Gene ontology analysis and connectivity mapping

The overlapping gene set was tested for enrichment of GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) terms using the GEne SeT AnaLysis Toolkit (Webgestalt) (Wang et al., 2013). A user-specific background gene set was used, which containing all expressed genes in all samples of NPCs with significant detection ($P < 0.05$). Gene ontology analysis may reveal biological processes implicated in the knockdown, and consequently events potentially associated with risk for schizophrenia. The overlapping gene list was also analysed by connectivity mapping using the Library of Integrated Network-based Cellular Signatures (LincsCloud) (http://apps.lincscloud.org/) (Vempati et al., 2014). At the time of analysis (Aug 2016), the database contained the expression profiles from

almost 80 human cell lines, including NPCs, following exposure to more than 20,000 drugs and genetic manipulation of over 20,000 genes. As a catalogue of gene expression response profiles, the platform allows the identification of drugs estimated to reverse a an input signature provided by the user based on empirical evidence contained in the database. The output provided includes compounds and the associated connectivity scores. A positive connectivity score is given to a drug that empirically elicited a transcriptomic signature in the cells in the database similarly to the user-input signature, whereas a negative score is given to a drug associated with potential to reverse the query signature. There has been some promising evidence for drug repositioning in other fields of medicine using connectivity mapping (Lamb, 2007). Although common risk mechanisms for schizophrenia (such as the 16% reduction in expression of NT5C2 associated with both risk variants in the DLPFC) are associated with small increase to disease susceptibility, it is possible that several risk mechanisms converge into similar changes to the global gene expression profile. Therefore, this analysis might provide clues about biological processes involved in schizophrenia and may expose drugs with potential to reverse them. The connectivity mapping results were filtered for the top 30 positive and top 30 negative connections in neural progenitor cells in order to keep only relevant compounds for the cellular model used in this study, but the top 400 connections are shown in Appendix 3.

5.4. Results

5.4.1. Validation of the siRNAs

In order to identify genes and genetic pathways affected by the downregulation of *NT5C2,* a transient silencing of this gene was performed using small interfering RNA (siRNA). Three siRNA sequences were initially tested for this purpose, and the samples treated with the two conditions that yielded the best knockdown were selected for global gene expression profiling using microarrays. First, transfection of the NPCs with the positive control reagent, BLOCK-iT™ Fluorescent Oligo, showed an incorporation of the red fluorescent oligonucleotides by 90.9% of cells (±0.02, CI 95%; n=4) (Figure 3). This is slightly higher than the transfection efficiency observed by Hill et al. (2012) for CTX0E03 (approximately 80%), which is a clonal population obtained from the same foetus as the CTX0E16 line (Pollock et al., 2006).

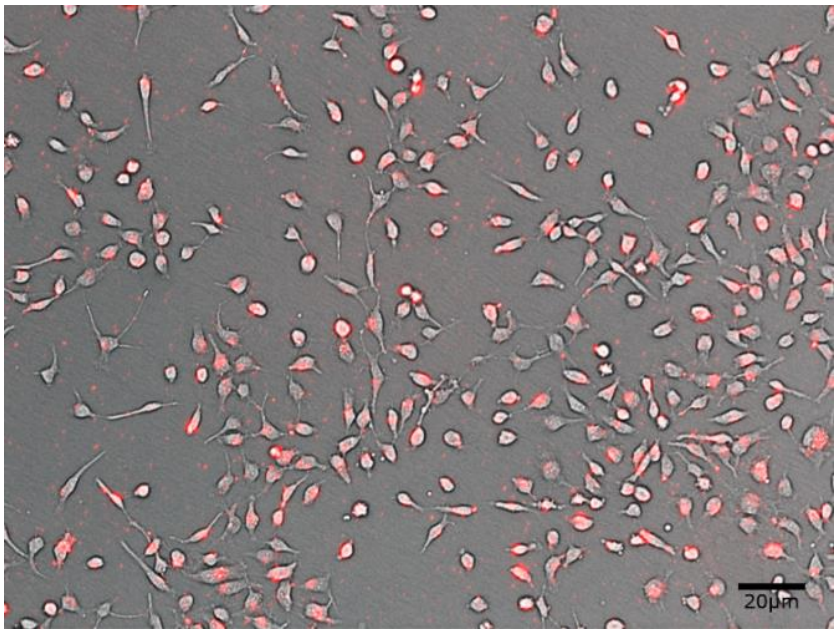


Figure 3. CTX0E16 transfected with *BLOCK-iT*™ fluorescent oligonucleotide (positive control of transfection). Cells were transfected using the N-TER™ Nanoparticle siRNA Transfection System. Efficiency of transfection was estimated to be 90.9% (±0.02, CI 95%) (n=4). Bright field images were overlaid with images captured using fluorescence and a 568 nm filter.

Expression of *NT5C2* was assessed using RT-qPCR in all samples (n=4) for estimation of the knockdown efficacy associated with the three siRNA sequences in RNA samples that would be submitted for sequencing. siRNAs A and B were the most efficient in reducing *NT5C2* transcripts in culture, yielding a reduction of 26.0% and 19.6%, as compared to the control (two-tailed t-tests, P=1.18E-03 and P=1.09E-03, respectively) (Figure 4). In order to keep exclusively relevant samples for the microarray analysis, samples submitted to siRNA C and one culture transfected with siRNA A were dropped due to poor silencing.

An independent immunocytochemistry assay using similar siRNA transfection conditions revealed the ability of siRNAs A and B to reduce NT5C2 at the protein level, as assessed by measures of Corrected Total Cell Fluorescence (CTCF) (Figure 5). siRNA A was associated with a reduction of 40% in NT5C2 immunostaining ($P=1x10^{-5}$), whereas siRNA B was associated with a reduction of 37% in NT5C2 labelling ($P=4x10^{-4}$). This indicated that changes in the global gene expression profile caused by the knockdown of *NT5C2* could be accounted for by reduced transcript and protein levels.

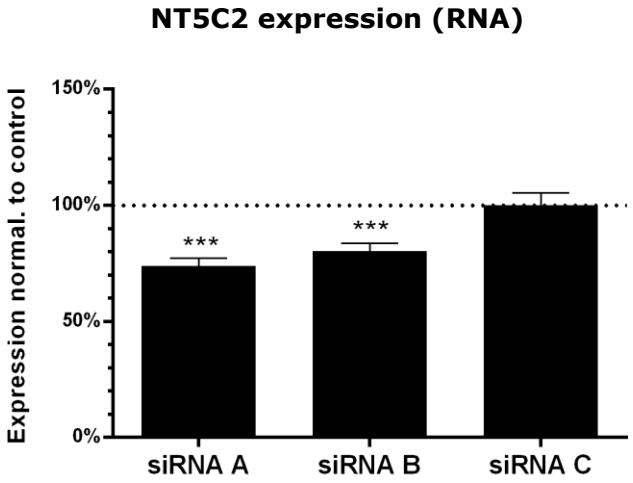
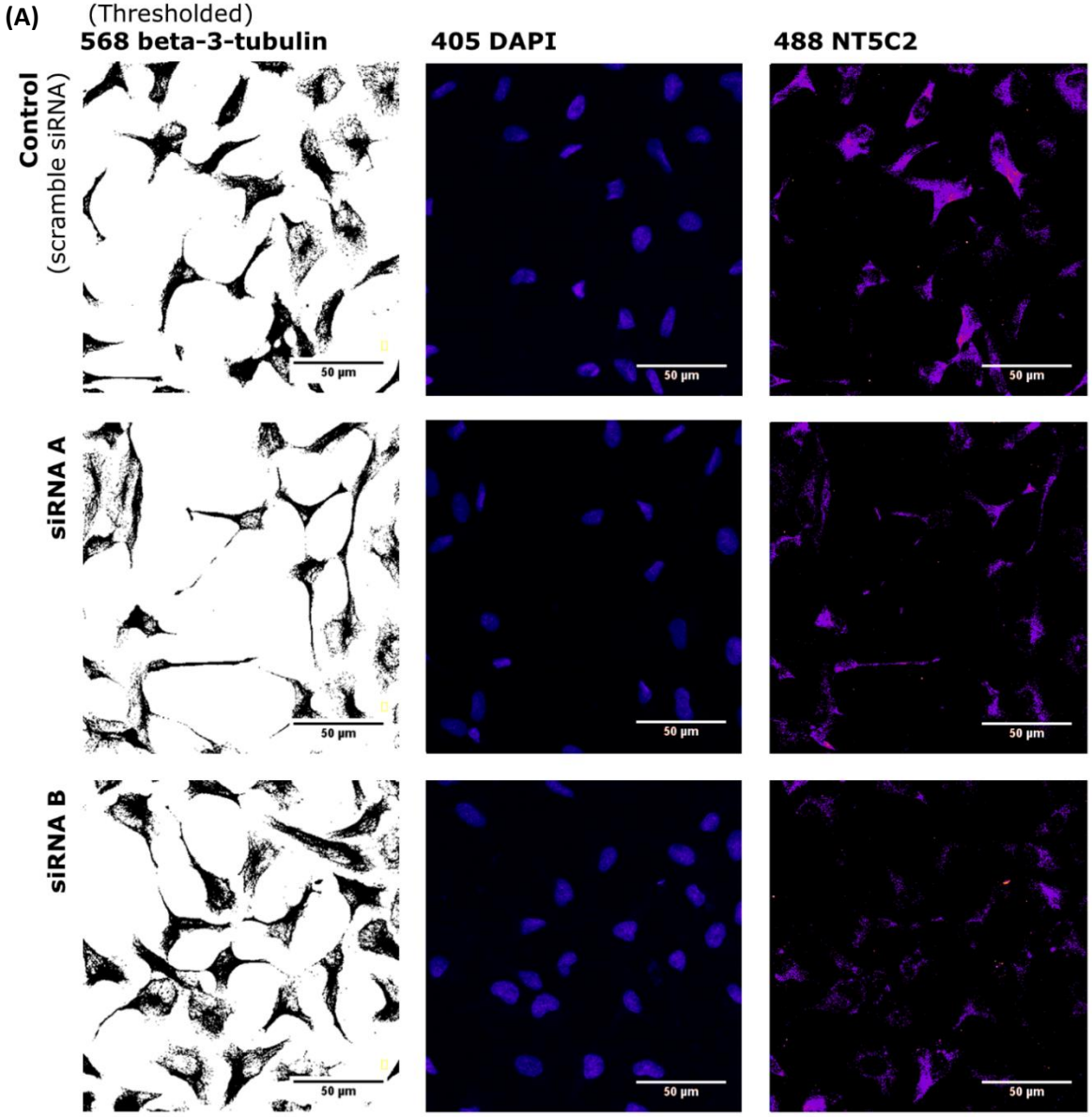


Figure 4. The effect of three siRNA sequences on *NT5C2* RNA expression, as measured by qPCR. The CTX0E16 cells were treated with 10 nM siRNAs. siRNA A was associated with a 26% reduction (n=3), and B with a 19.6% reduction on overall NT5C2 expression (n=4), as compared to expression at the control conditions (dotted line). *** $P \leq 0.001$, two-tailed t-test of Pfaffl normalised values for *NT5C2* expression.



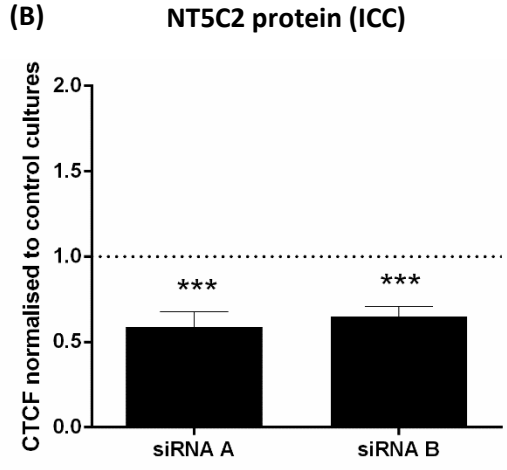


Figure 5. *NT5C2* siRNA conditions are associated with reduced protein levels, as measured by immunocytochemistry. Reduced immunostaining of NT5C2 was observed after 72-hours transfection with siRNAs A (40% reduction) and B (37% reduction) in CTX0E16 NPCs. **(A)** Representative immunocytochemistry images where ROIs can be visualised in the three channels (morphological marker Beta-3-tub, DAPI and NT5C2) for the three siRNA conditions (A, B and control). Scale bar represents 50µm. **(B)** Corrected Total Cell Fluorescence (CTCF) values were measured for approximately 320 cells per condition (n=4), normalised against the control for estimates of the knockdown, and given in percentage. *** P<0.001, one-way ANOVA, Dunnet-corrected.

5.4.2. Microarray results and validation

Cultures in which *NT5C2* expression was transiently knockdown were submitted for global gene expression profiling. More than 47,000 probes of the Illumina HT12 v4 chip, representing more than 20,000 genes, were used to investigate changes in global gene expression associated with this putative schizophrenia risk mechanism. A total of 12,774 genes were found to be expressed in all samples and conditions with a detection threshold $p<0.05$. From these, 881 were differentially expressed in association with siRNA A ($p<0.05$, uncorrected), and 741 in association with siRNA B ($p<0.05$, uncorrected), as compared to the control conditions (Figure 6). Differential expression was validated using qPCR for some genes (Table 2) in the overlapping set of 74 genes (Table 3), for which significant expression changes were shared by two siRNA sequences targeting *NT5C2.* This overlapping gene set is unexpected to be observed by chance, according to a hypergeometric probability ($P=5.92x10^{-5}$), suggesting that a substantial proportion of these were genuinely associated with the knockdown. It is important to note, however, that the hypergeometric probability calculated is likely inflated since the present experiment deviates from the assumption of independent controls (as shared controls were used).

Four genes from the overlapping gene set were chosen for qPCR validation based on size of effect and statistical confidence (Table 2). The ability to detect these changes by qPCR was tested in cDNA synthesised from the same RNA samples used for global gene expression profiling. Expression of *PSMC4* (proteasome 26S subunit, ATPase 4) and *HNRNPA1* (heterogeneous nuclear ribonucleoprotein A1) was found to be significantly reduced according to both methods, in both siRNA conditions. Expression of *ATG4B* (autophagy related 4B cysteine peptidase), similarly, was decreased upon knockdown, but this difference was not significant upon treatment with siRNA A, according to qPCR analysis. Expression of *TBCA* (tubulin-specific chaperone A), however, was found to be increased as per microarray assessment, but decreased by qPCR measurement. *TBCA* was removed from further analysis, and the possible reasons for this discrepancy are presented in the Discussion.

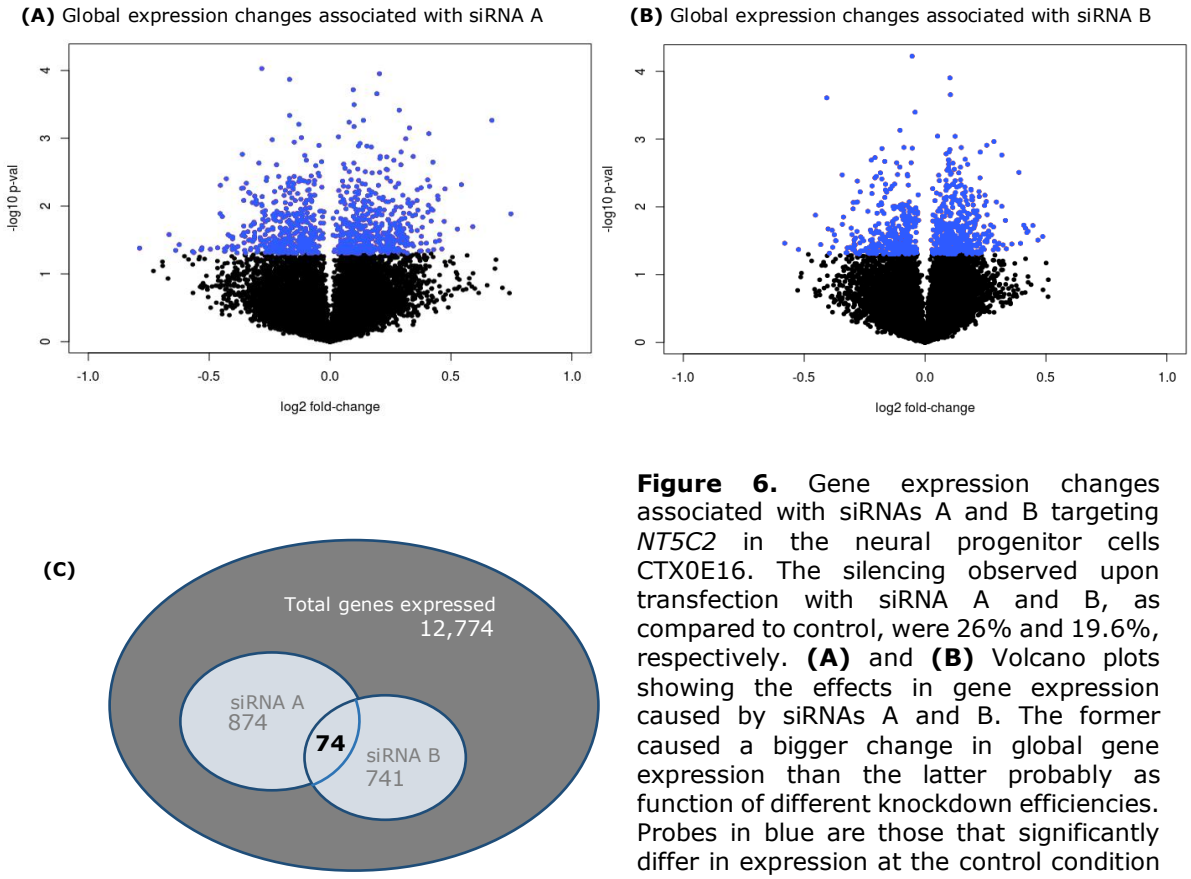


Figure 6. Gene expression changes associated with siRNAs A and B targeting *NT5C2* in the neural progenitor cells CTX0E16. The silencing observed upon transfection with siRNA A and B, as compared to control, were 26% and 19.6%, respectively. **(A)** and **(B)** Volcano plots showing the effects in gene expression caused by siRNAs A and B. The former caused a bigger change in global gene expression than the latter probably as function of different knockdown efficiencies. Probes in blue are those that significantly differ in expression at the control condition (P<0.05, uncorrected). **(C)** Venn diagram representing the overlapping gene sets in the context of the global gene expression. From the overlapping gene set of 74 genes, only 65 have a known function.

Table 2. RT-qPCR validation of four gene expression changes associated with both NT5C2 knockdown conditions.

Gene	Fold-change array		P-value array		Fold-change qPCR		P-value qPCR	
	siRNA A	siRNA B	siRNA A	siRNA B	siRNA A	siRNA B	siRNA A	siRNA B
ATG4B	0.79455	0.79126	0.00645	0.01411	0.73033	0.89265	0.18794	0.04843
PSMC4	0.69212	0.79923	0.04311	0.04717	0.63005	0.86497	0.00570	0.01361
HNRNPA1	0.74267	0.80920	0.00394	0.02055	0.62572	0.74492	0.00444	0.00001
*TBCA**	1.32271	1.21001	0.00409	0.02579	0.82561	0.90537	0.00805	0.10001

*Discordant gene, removed from further analysis.

Table 3. Genes differentially expressed (at P < 0.05) in association with the knockdown of NT5C2.

Gene symbol	siRNA A		siRNA B	
	P-value	Fold-change	P-value	Fold-change
FZD7	0.002	0.934	0.183	0.946
RPL6	0.002	0.858	0.558	0.938
PTPLAD1	0.003	0.967	0.043	0.934
HNRNPA1	0.004	0.743	0.021	0.809
TPM1	0.004	0.833	0.181	0.889
CNIH	0.005	0.868	0.401	0.857
LOC729799	0.005	0.977	0.014	0.936
ATG4B	0.006	0.795	0.014	0.791
ILF3	0.006	0.9	0.271	0.925
DSTN	0.007	0.944	0.035	0.949
LARP1B	0.014	0.83	0.87	0.982
MEST	0.014	0.905	0.034	0.922
MSMO1	0.014	0.849	0.006	0.83
C21ORF66	0.016	0.951	0.862	0.995
NSUN5	0.019	0.913	0.016	0.953
NFX1	0.021	0.867	0.012	0.884
LOC440589	0.022	0.817	0.034	0.893
CLEC2D	0.024	0.929	0.809	0.991
FBXO38	0.025	0.945	0.026	0.908
GTF2I	0.027	0.805	0.049	0.861
ACRC	0.032	0.935	0.027	0.936
SLC15A4	0.032	0.825	0.646	0.977
EZH2	0.035	0.916	0.408	0.945
DLGAP5	0.042	0.797	0.31	0.858
PSMC4	0.043	0.692	0.047	0.799
CEP152	0.047	0.901	0.282	0.924
RPL22	0.048	0.831	0.018	0.953
ACTG1	0.049	0.785	0.031	0.774
RAE1	0.049	0.948	0.027	0.936
WDR4	0.049	0.967	0.029	0.819
CYP27A1	0.002	1.071	0.015	1.035
NUDC	0.003	1.106	0.007	1.168
*TBCA	0.004	1.323	0.026	1.21
DACT3	0.005	1.097	0.045	1.073
PRDX1	0.005	1.298	0.03	1.194
ATP6AP2	0.006	1.284	0.035	1.282
KBTBD7	0.007	1.232	0.018	1.059
CEND1	0.008	1.063	0.018	1.096
PKM	0.008	1.197	0.033	1.194
AHNAK	0.011	1.343	0.118	1.29
ATXN1	0.011	1.153	0.044	1.116
RABGAP1L	0.012	1.019	0.031	1.017
ACTN1	0.013	1.142	0.041	1.175
RHBDD2	0.015	1.229	0.023	1.115
CMTM7	0.016	1.181	0.035	1.091
PORCN	0.018	1.039	0.032	1.059
HMGN1	0.020	1.259	0.006	1.167
LOC150568	0.021	1.086	0.026	1.073
TM9SF1	0.021	1.139	0.024	1.035
PDP2	0.024	1.116	0.121	1.037
P4HA2	0.025	1.08	0.007	1.078
FOXO1	0.026	1.168	0.026	1.122
PLCE1	0.028	1.151	0.005	1.082
RDX	0.028	1.299	0.042	1.233
CNOT1	0.031	1.091	0.032	1.187
ECD	0.031	1.117	0.098	1.129
MMS19	0.031	1.24	0.04	1.18
TPM2	0.032	1.039	0.006	1.062
C20ORF199	0.033	1.197	0.013	1.136
HS.363526	0.034	1.141	0.005	1.068
POLDIP3	0.035	1.236	0.424	1.108
DNAJB12	0.037	1.051	0.012	1.046
LOC646345	0.037	1.253	0.046	1.132
C17ORF97	0.038	1.09	0.014	1.111
CMTM4	0.038	1.242	0.004	1.04
SLC38A9	0.038	1.064	0.002	1.062

Table 3: continued.

TargetID	siRNA A		siRNA B	
	P-value	Fold-change	P-value	Fold-change
MAPK1	0.04	1.103	0.025	1.075
ATP5D	0.041	1.176	0.008	1.125
C1ORF123	0.041	1.059	0.043	1.124
TGFA	0.042	1.215	0.016	1.154
VPS26B	0.042	1.215	0.027	1.152
FBXO17	0.044	1.065	0.048	1.032
FAM53C	0.046	1.235	0.001	1.193
APPBP2	0.048	1.069	0.012	1.067

Gene names in blue represent downregulation; in red, upregulation. Highlighted genes were not recognised as part of genetic pathways. *Gene excluded from analysis, as directionality in expression measured by qPCR was different from microarray assessment.

5.4.3. Gene ontology analysis

In order to explore the biological mechanisms associated with the downregulation of *NT5C2*, gene ontology analysis was applied to the overlapping gene set. The GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases are major bioinformatic catalogues of genes and associated biological functions. From the full overlapping gene set of 74 genes differentially regulated by the knockdown, 65 genes of known function were assessed for enrichment for particular GO and KEGG terms. Several expression, localisation and gene-interaction modules functionally connect the genes on the overlapping gene set, supporting their association (Figure 7). The resulting terms showed redundancy (i.e. similar terms referring to the same subset of genes) and were thus grouped into similar categories. The GO terms significantly enriched within the differentially expressed genes associated with the knockdown (Table 4), at a false discovery rate cut-off (FDR) q<0.05, included cellular component terms related to:

- **thin muscle** (GO:0005862, FDR=0.0174 / GO:0005865, FDR=0.0174),
- **contractile fibers** (GO:0030016, FDR=0.0290 / GO:0043292 FDR=0.0325),
- **pseudopodium formation** (GO:0031143, FDR=0.0290).

Terms linked with the knockdown, with confidence FDR<0.1 included:

- **stress response** (GO:0070302, P=0.0005 / GO:0032872, P=0.0005),
- **nuclear transport** (GO:0051169, P=0.0011 / GO:0006913, P=0.0011),
- **activation of the mitogen-activated protein kinases (MAPK)** (GO:0043408, P=0.0004 / GO:0032872, P=0.0005 / GO:0000187, P=0.0011 / GO:0000165, P=0.001).

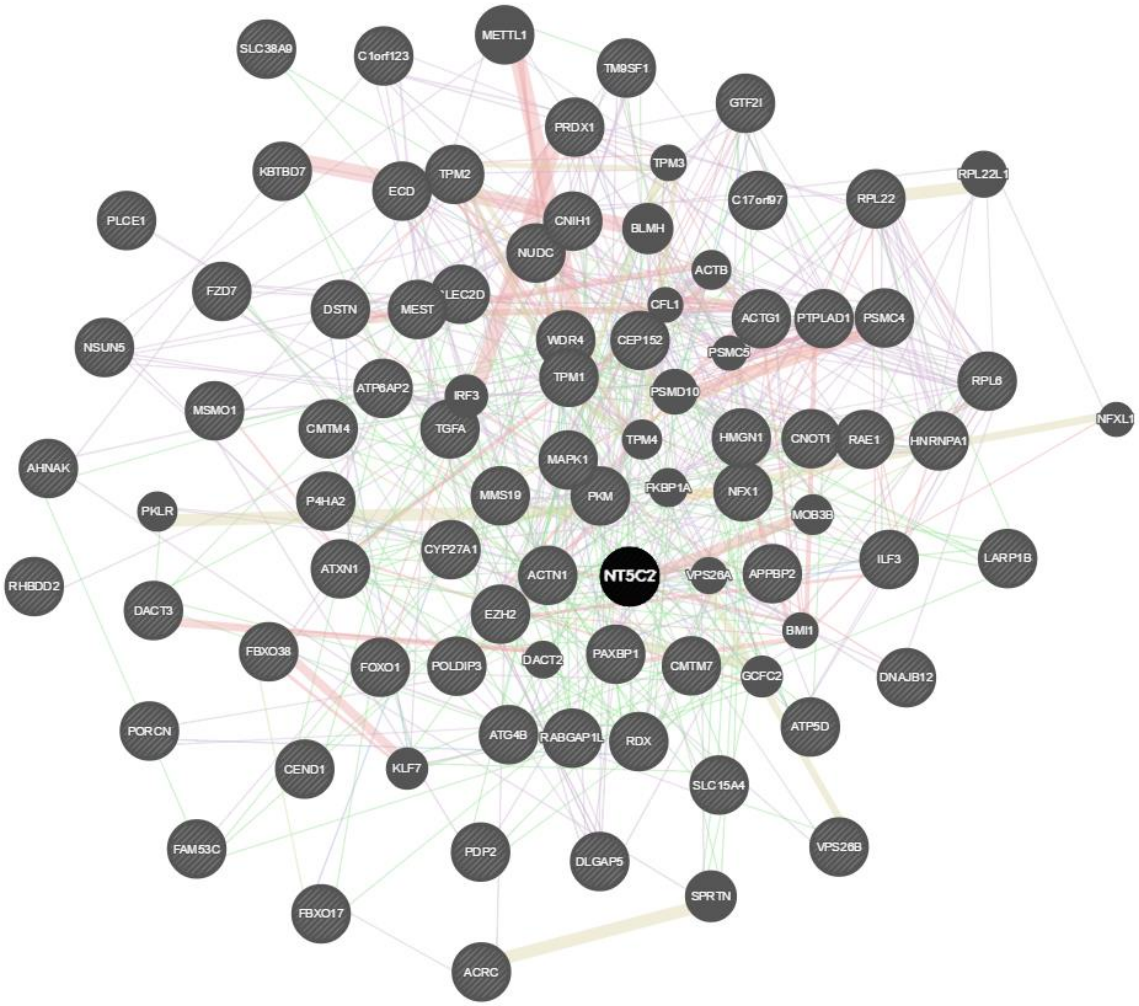


Figure 7. Genetic network associated with the knockdown of NT5C2. Modules shown are of co-expression (purple, 64.7% of connections), physical interactions (pink, 23.1%), co-localisation (blue, 4.0%), genetic interactions (green, 3.4%), shared protein domains (yellow, 2.6%). *NT5C2* is found well connected to members of this network in the very centre. Created with GeneMania (www.genemania.org).

Table 4. GO terms enriched amongst differentially expressed genes following *NT5C2* knockdown in NPCs.

Cellular components	**GO:0005862** **GO:0005865**	**Muscle thin filament tropomyosin** **Striated muscle thin filament** C=4;O=2;E=0.03;R=73.01;rawP=0.0003;**FDR =0.0174**		
	TPM1	tropomyosin 1 (alpha)	7168	ENSG0000014041
	TPM2	tropomyosin 2 (beta)	7169	ENSG0000019846
	GO:0031143	**Pseudopodium** C=7;O=2;E=0.05;R=41.72;rawP=0.0009;**FDR=0.0290**		
	ACTN1	actinin, alpha 1	87	ENSG0000007211
	MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
	GO:0030016 **GO:0043292**	**Myofibril** C=65;O=4;E=0.45;R=8.99;rawP=0.0010;**FDR=0.0290** **Contractile fiber** C=72;O=4;E=0.49;R=8.11;rawP=0.0014;**FDR=0.0325**		
	ACTN1	actinin, alpha 1	87	ENSG0000007211
	TPM1	tropomyosin 1 (alpha)	7168	ENSG0000014041
	TPM2	tropomyosin 2 (beta)	7169	ENSG0000019846
	ACTG1	actin, gamma 1	71	ENSG0000018400
	GO:0030017	**Sarcomere** C=50;O=3;E=0.34;R=8.76;rawP=0.0048;**FDR=0.0928**		
	ACTN1	actinin, alpha 1	87	ENSG0000007211
	TPM1	tropomyosin 1 (alpha)	7168	ENSG0000014041
	TPM2	tropomyosin 2 (beta)	7169	ENSG0000019846
Biological processes	**GO:0043408** **GO:0000165**	**Regulation of MAPK cascade** C=253;O=8;E=1.78;R=4.49;rawP=0.0004;**FDR=0.0979** **MAPK cascade** C=296;O=8;E=2.08;R=3.84;rawP=0.0010;**FDR=0.0979**		
	PTPLAD1	protein tyrosine phosphatase-like A domain 1	51495	ENSG0000007469
	FOXO1	forkhead box O1	2308	ENSG0000015090
	MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
	TGFA	transforming growth factor, alpha	7039	ENSG0000016323
	FZD7	frizzled family receptor 7	8324	ENSG0000015576
	PRDX1	peroxiredoxin 1	5052	ENSG0000011745
	PLCE1	phospholipase C, epsilon 1	51196	ENSG0000013819
	ATP6AP2	ATPase, H+ transporting, lysosomal accessory p2	10159	ENSG0000018222
	GO:0032872 **GO:0070302**	**Regulation of stress-activated MAPK** **Regulation stress-activated kinase**		
	PTPLAD1	protein tyrosine phosphatase-like A domain 1	51495	ENSG0000007469
	FOXO1	forkhead box O1	2308	ENSG0000015090
	MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
	FZD7	frizzled family receptor 7	8324	ENSG0000015576
	PRDX1	peroxiredoxin 1	5052	ENSG0000011745
	GO:0000187	**Activation of MAPK activity** C=66;O=4;E=0.46;R=8.61;rawP=0.0011;**FDR=0.0979**		
	PTPLAD1	protein tyrosine phosphatase-like A domain 1	51495	ENSG0000007469
	MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
	TGFA	transforming growth factor, alpha	7039	ENSG0000016323
	PLCE1	phospholipase C, epsilon 1	51196	ENSG0000013819
	GO:0006913 **GO:0051169**	**Nucleocytoplasmic transport** C=230;O=7;E=1.62;R=4.33;rawP=0.0011;**FDR=0.0979** **Nuclear transport** C=232;O=7;E=1.63;R=4.29;rawP=0.0011;**FDR=0.0979**		
	ATXN1	ataxin 1	6310	ENSG0000012478
	HNRNPA1	heterogeneous nuclear ribonucleoprotein A1	3178	ENSG0000013548
	DACT3	dapper, antagonist of beta-catenin, homolog 3	147906	ENSG0000019738
	MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
	RAE1	RAE1 RNA export 1 homolog (S. pombe)	8480	ENSG0000010114
	FZD7	frizzled family receptor 7	8324	ENSG0000015576
	PRDX1	peroxiredoxin 1	5052	ENSG0000011745

* **Legend**: number of reference genes in the category (C), number of genes in the gene set and also in category (O), expected number in the category (E), Ratio of enrichment (R), p-value from hypergeometric test (rawP), and p-value FDR-adjusted (FDR). For each gene, its description, Entrez ID and Ensembl ID are given on the same row.

The KEGG terms significantly enriched within the differentially expressed genes associated with the knockdown (Table 5), at a FDR<0.05, included cardiomyopathy related pathways (5410, FDR=0.0149 / 5414, FDR=0.0149), as well as adherens junctions (ID 4520, FDR=0.0376), regulation of actin cytoskeleton (ID 4810, FDR=0.0432) and prostate cancer (ID 5215, FDR=0.0432). Enriched terms at FDR<0.1 included additional cardiovascular (5412 and 4260, FDR=0.0697) and cancer-related modules (5200, 5211 and 5214, FDR=0.0814).

An enrichment of regulatory pathways related to the mitogen-activated protein kinases (MAPK) was observed amongst differentially expressed genes (FDR<0.1). Therefore, activation of MAPK signalling was investigated using western blotting to provide measures of increased activation of ERK1/2 (extracellular signal-regulated kinases), members of the MAPK cascade, in association with the knockdown. Activation of ERK1/2 (p42/p44) was measured by quantification of phosphorylation at residues Thr202/Tyr204 (Figure 8A). Although a mean increased phosphorylation of MAPK-ERK1/2 was observed in comparison to the control (siRNA A: 60% increased phosphorylation; siRNA B: 25% increased phosphorylation; P>0.05, n=4), this difference was not statistically significant. Activation of the AMP-activated protein kinase (AMPK), a key-regulator kinase of cellular metabolism and nucleotide pools, was also investigated. This analysis provided evidence of increased phosphorylation of AMPK-alpha at residue Thr172, associated with the knockdown (Figure 8B) (siRNA A: 40% increased phosphorylation; siRNA B: 65% increased phosphorylation; P=0.03, n=4). These results support a role for AMPK signalling as a potentially downstream cellular effect associated with the downregulation of *NT5C2*.

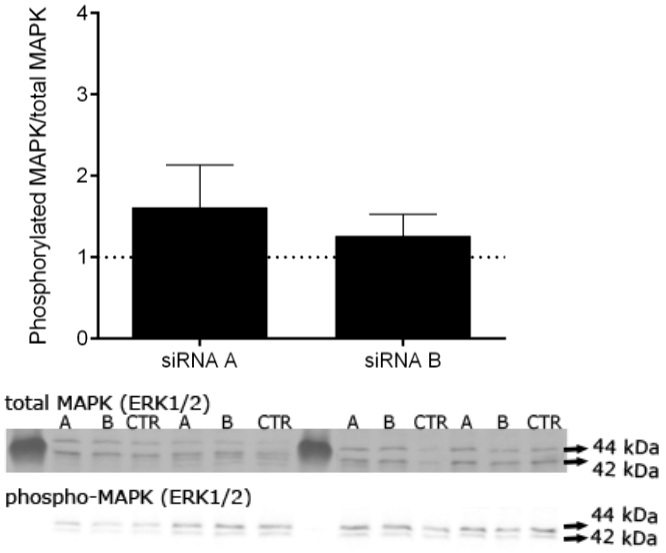
Table 5. KEGG terms enriched amongst differentially expressed genes associated with the *NT5C2* knockdown.

KEGG pathways

5410	**Hypertrophic cardiomyopathy** (HCM)		
5414	**Dilated cardiomyopathy** C=33;O=3;E=0.22;R=13.70;rawP=0.0013;**FDR=0.0149**		
TPM1	tropomyosin 1 (alpha)	7168	ENSG0000014041
ACTG1	actin, gamma 1	71	ENSG0000018400
TPM2	tropomyosin 2 (beta)	7169	ENSG0000019846
4520	**Adherens junction** C=52;O=3;E=0.35;R=8.69;rawP=0.0049;**FDR=0.0376**		
ACTN1	actinin, alpha 1	87	ENSG0000007211
ACTG1	actin, gamma 1	71	ENSG0000018400
MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
4810	**Regulation of actin**		
RDX	Radixin	5962	ENSG0000013771
ACTN1	actinin, alpha 1	87	ENSG0000007211
ACTG1	actin, gamma 1	71	ENSG0000018400
MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
5215	**Prostate cancer** C=65;O=3;E=0.43;R=6.95;rawP=0.0091;**FDR=0.0432**		
FOXO1	forkhead box O1	2308	ENSG0000015090
MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
TGFA	transforming growth factor, alpha	7039	ENSG0000016323
5412	**Arrhythmogenic ventricular cardiomyopathy**		
ACTN1	actinin, alpha 1	87	ENSG0000007211
ACTG1	actin, gamma 1	71	ENSG0000018400
4260	**Cardiac muscle contraction** C=34;O=2;E=0.23;R=8.86;rawP=0.0212;**FDR=0.0697**		
TPM1	tropomyosin 1 (alpha)	7168	ENSG0000014041
TPM2	tropomyosin 2 (beta)	7169	ENSG0000019846
5200	**Pathways in cancer** C=210;O=4;E=1.39;R=2.87;rawP=0.0505;**FDR=0.0814**		
FOXO1	forkhead box O1	2308	ENSG0000015090
MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
TGFA	transforming growth factor, alpha	7039	ENSG0000016323
FZD7	frizzled family receptor 7	8324	ENSG0000015576
5211	**Renal cell carcinoma** C=52;O=2;E=0.35;R=5.79;rawP=0.0465;**FDR=0.0814**		
5214	**Glioma** C=45;O=2;E=0.30;R=6.70;rawP=0.0358;**FDR=0.0814**		
MAPK1	mitogen-activated protein kinase 1	5594	ENSG0000010003
TGFA	transforming growth factor, alpha	7039	ENSG0000016323

* **Legend**: number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p-value from hypergeometric test (rawP), and p value FDR-adjusted (FDR). For each gene, its description, Entrez ID and Ensembl ID are given on the same row.

(A) **Phospho-ERK1/2 (MAPK-p42/p44) (Thr202/Tyr204)**



(B) **Phospho-AMPK-alpha (Thr172)**

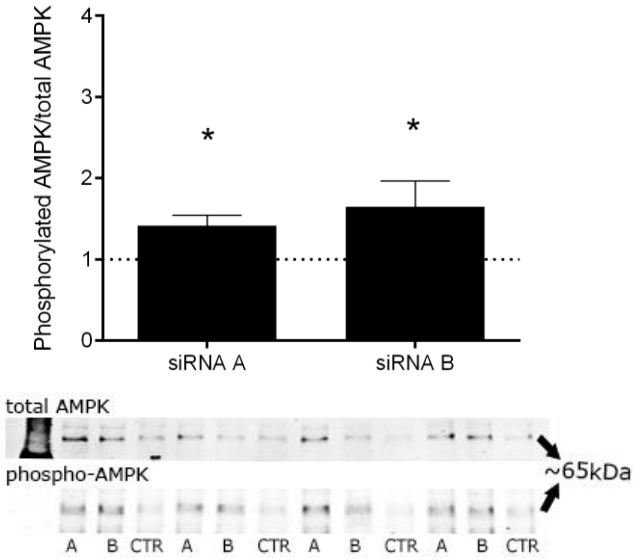


Figure 8. Influence of *NT5C2* knockdown on activation of MAPK and AMPK cascades. The transient knockdown of *NT5C2* was associated with significant activation of AMPK signalling, and with increased mean activation of MAPK signalling, although the latter was not statistically significant. **(A)** Phosphorylation of MAPK (mitogen-activated protein kinases) at Thr202/Tyr204 of ERK1/2 (p42/p44). Increased mean activation of the MAPK signalling cascade was observed in association with the knockdown, but this event was not statistically significant. Mean phosphorylation was increased by 60% and 25% (n=4), as elicited by transfection with siRNAs A and B, respectively, when compared to the phosphorylation levels in the control condition (dotted line). **(B)** Phosphorylation of AMPK (AMP-activated protein kinase) at residue Thr172 was found increased upon knockdown of *NT5C2* with two siRNA sequences. Densitometry quantification (top) of Western blot bands (bottom; A=siRNA A, B=siRNA B; CTR=control) represent phosphorylation at Thr172 normalised by total AMPK, adjusted to basal AMPK activation in the control (dotted line). siRNA A: 40% increased phosphorylation; siRNA B: 65% increased phosphorylation; *P<0.05, non-parametric analysis of variance, n=4.

5.4.4. Connectivity mapping

Connectivity mapping was used to identify genes and chemicals that could elicit or reverse the effects on gene expression associated with the knockdown of *NT5C2* in the neural progenitor cells. The top 30 positive and negatively correlated connections are shown on Table 6, but a more comprehensive list of the top 400 connections is listed on Appendix 3. Interestingly, present amongst the top chemical compounds estimated to counteract the knockdown (highlighted in blue) are the antipsychotics ziprasidone and fluphenazine, atypical and typical drugs, respectively. Other chemicals in this list include drugs that could potentially be repurposed for schizophrenia treatment, such as anandamine, an endogenous cannabinoid (De Petrocellis et al., 1998); eplerenone, a steroidal drug used for protection against cardiovascular disease (Davis and Nappi, 2003); hippeastrine, an alkaloid compound active against tumours and angiogenesis (Cao et al., 2013); and tomelukast, used to block inflammatory response in asthma (Berger, 1999).

Amongst top compounds expected to mimic the global expression signature elicited by the knockdown of *NT5C2* (highlighted in red) include proscillaridin, an inhibitor of the Na^+/K^+ ATPase pump used for treatment of heart conditions and known to trigger cell death in glioblastoma cells (Denicolai et al., 2014); ritodrine, a drug used to delay premature labour, recently described as teratogenic (Boga Pekmezekmek et al., 2015); strophanthidin, an inhibitor of the Na^+/K^+ ATPase pump shown to increase glutamate-evoked intracellular Ca^{+2} signalling (Song et al., 2014); and menadione, a drug that induces Parkinsonian symptoms (Janda et al., 2015).

Table 6. Top connectivity mapping results for the *NT5C2* knockdown signature in NPCs.

Overall rank	Perturbant	Class	Cells	Exposure	Dose	Connectivity score
476221	ziprasidone	Chemical	NPC	24 h	10 µM	-0.4975
476187	CHIR-99021	Chemical	NPC	6 h	10 µM	-0.4862
476110	OXCT1	Gene	NPC	96 h	1.5 µL	-0.4729
476096	TCIRG1	Gene	NPC	96 h	1.5 µL	-0.4710
476005	fluphenazine	Chemical	NPC	24 h	10 µM	-0.4622
475935	APHA-compound-8	Chemical	NPC	24 h	10 µM	-0.4580
475905	NCAPD2	Gene	NPC	96 h	1.5 µL	-0.4564
475866	anandamide	Chemical	NPC	24 h	10 µM	-0.4538
475853	TIAM1	Gene	NPC	96 h	1.5 µL	-0.4532
475821	PAF1	Gene	NPC	96 h	1.5 µL	-0.4515
475811	UBAP1	Gene	NPC	96 h	1.5 µL	-0.4508
475812	PHF15	Gene	NPC	96 h	1.5 µL	-0.4508
475808	BRD-K31302860	Chemical	NPC	24 h	10 µM	-0.4507
475783	TCFL5	Gene	NPC	96 h	1.5 µL	-0.4494
475764	CIAPIN1	Gene	NPC	96 h	1.5 µL	-0.4486
475719	GTPBP8	Gene	NPC	96 h	1.5 µL	-0.4466
475720	DNAJB12	Gene	NPC	96 h	1.5 µL	-0.4466
475616	BRD-K69569876	Chemical	NPC	24 h	10 µM	-0.4431
475537	AP1B1	Gene	NPC	96 h	1.5 µL	-0.4401
475518	BRD-K33164466	Chemical	NPC	24 h	10 µM	-0.4397
475375	PUF60	Gene	NPC	96 h	1.5 µL	-0.4352
475362	KDM3A	Gene	NPC	96 h	1.5 µL	-0.4348
475321	STK10	Gene	NPC	96 h	1.5 µL	-0.4337
475262	EPHB4	Gene	NPC	96 h	1.5 µL	-0.4324
475058	eplerenone	Chemical	NPC	24 h	10 µM	-0.4276
475030	hippeastrine	Chemical	NPC	24 h	10 µM	-0.4271
474986	tomelukast	Chemical	NPC	24 h	10 µM	-0.4263
474928	TIAM1	Gene.cgs	NPC	96 h	1.5 µL	-0.4254
474857	TRIB3	Gene	NPC	96 h	1.5 µL	-0.4241
474832	CLASRP	Gene	NPC	96 h	1.5 µL	-0.4237
113	F-1061-0166	Chemical	NPC	24 h	10 µM	0.4953
140	BRD-K66902379	Chemical	NPC	24 h	10 µM	0.4917
146	ritodrine	Chemical	NPC	24 h	10 µM	0.4903
181	BRD-K56994829	Chemical	NPC	6 h	10 µM	0.4865
198	WZ-3105	Chemical	NPC	24 h	10 µM	0.4835
233	PEPD	Gene.cgs	NPC	96 h	1.5 µL	0.4801
275	MD-II-008-P	Chemical	NPC	24 h	10 µM	0.4766
325	ER-27319	Chemical	NPC	24 h	10 µM	0.4728
345	SB-431542	Chemical	NPC	24 h	10 µM	0.4718
358	A-443644	Chemical	NPC	24 h	10 µM	0.4711
453	GSK-1070916	Chemical	NPC	24 h	10 µM	0.4655
509	BAZ1B	Gene	NPC	96 h	1.5 µL	0.4623
514	PQ-401	Chemical	NPC	24 h	10 µM	0.4622
562	AURKA	Gene	NPC	96 h	1.5 µL	0.4604
566	CCDC90A	Gene	NPC	96 h	1.5 µL	0.4603
619	AS-601245	Chemical	NPC	24 h	10 µM	0.4584
680	FZD5	Gene	NPC	96 h	1.5 µL	0.4566
738	BRD-K31290914	Chemical	NPC	24 h	10 µM	0.4545
769	BRD-K31553917	Chemical	NPC	24 h	10 µM	0.4537
789	proscillaridin	Chemical	NPC	24 h	10 µM	0.4528
793	SLC35A3	Gene	NPC	96 h	1.5 µL	0.4527
844	strophanthidin	Chemical	NPC	24 h	10 µM	0.4514
862	EPB41L4B	Gene	NPC	96 h	1.5 µL	0.4510
875	HDAC3-selective	Chemical	NPC	24 h	10 µM	0.4506
900	BRD-K92993117	Chemical	NPC	24 h	10 µM	0.4499
915	RRP8	Gene	NPC	96 h	1.5 µL	0.4495
929	LY-303511	Chemical	NPC	24 h	10 µM	0.4492
950	MAPK1IP1L	Gene	NPC	96 h	1.5 µL	0.4487
955	STMN1	Gene	NPC	96 h	1.5 µL	0.4486
998	menadione	Chemical	NPC	24 h	10 µM	0.4475

*The top 400 connections is listed on the Appendix 3. Compounds predicted to reverse the knockdown signature are shown in blue, and those predicted to mimic the knockdown signature are shown in red.

5.5. Discussion

Chromosome 10q24 is currently the third most significantly associated locus to emerge from large-scale GWAS of schizophrenia (Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Schizophrenia Working Group of the PGC, 2014). As observed in Chapter 2, the schizophrenia risk alleles are associated with reduced expression of *NT5C2* in several brain regions. Reproducing this putative risk mechanism in a relevant cellular model was intended to expose downstream cellular pathways. The CTX0E16 cells have previously been shown to differentiate into action potential-firing, cortical glutamatergic cells (Anderson et al., 2015), substantiating their choice for this study.

Following transfection with the Trilencer 27-mer siRNA sequences specific for *NT5C2*, siRNA A showed the highest *NT5C2* silencing efficacy at the transcript and protein levels (26% and 40%, respectively), followed by siRNA B (19.6% and 37%) and siRNA C (where no silencing was observed) (Figures 4 and 5). Overall NT5C2 expression at the RNA level was assessed by probing exons common to all transcripts using qPCR, which would capture the expression of the novel variants detected in Chapter 3. The increased knockdown efficiency of siRNA A as compared to that of siRNA B was noted in the microarray results, in which the latter caused the differential regulation of 741 genes, while the former impacted expression of 874 genes (a difference of 133 genes). Nonetheless, the overlapping gene set of 74 nominally significant genes, for which directionality of effect was concordant upon transfection with two independent siRNAs, was found to be more than expected by chance, according to a hypergeometric probability. Such study design, as performed by Hill et al. (2012), potentially reduces off-target effects associated with individual siRNA sequences.

From the overlapping gene set, four were chosen for validation by RT-qPCR. As microarray probe intensities can be affected by complementarity and fluorescence artefacts, qPCR was used to confirm differential expression of selected genes. Accordingly, expression of *ATG4B* (an autophagy related protein), *PSCM4*

(proteasome ATPase subunit) and *HNRNPA1* (heterogeneous nuclear ribonucleoprotein A1) was found decreased during the knockdown of *NT5C2*, as measured by microarray probe intensities and qPCR, supporting the microarray findings (although the decrease in *ATG4B* expression elicited by siRNA A treatment did not reach statistical confidence in the qPCR validation). However, assessment of *TBCA* (tubulin folding cofactor A) expression in association with the knockdown produced discrepant results by qPCR and microarray, and therefore the gene was removed from further analysis (and, moreover, it was not attributed to any GO or KEGG terms). In general terms, microarrays provide a high-throughput workflow that is more affordable than RNA-seq, for instance, but background noise and eventual lack of reproducibility associated with microarray experiments raise a pertinent issue for molecular biologists, who have to find a balance between cost, and cohort and effect sizes (Jaksik et al., 2015). Microarrays provide a powerful means for global gene expression analysis in large sample sizes, where noise in individual probes in specific samples become diluted, but for smaller sample sizes, RNA-seq is more likely to provide accurate measures due to its increased sensitivity. The lack of specificity in microarrays, therefore, is likely the reason for the discrepancy observed. The single probe tagging *TBCA* on the Illumina HumanHT-12 v4 BeadChip was found to match other two pre-mRNAs in the genome, with 100% and 98% identity upon BLAST alignment. Moreover, *NT5C2* is also not amongst the differentially regulated genes upon siRNA treatments, although the knockdown was highly significant by qPCR. As observed by microarray probe intensities, the reduction of *NT5C2* expression elicited by siRNAs A and B was estimated to be 5% ($P=0.10$) and 9% ($P=0.41$), contrary to the highly significant 26% and 20% reduction associated with these siRNAs, as compared to the control, when assessed by qPCR ($P \leq 0.001$ for both comparisons).

The overlapping gene set, in which the effect on gene expression was shared between two independent siRNA sequences, was then submitted to gene ontology analysis and connectivity mapping methods. KEGG terms associated with the actin cytoskeleton and adherens junctions were significantly over-represented in the

overlapping gene set associated with *NT5C2* knockdown. Similarly, knockdown of the putative schizophrenia susceptibility gene *ZNF804A* was found to alter expression of genes involved in cellular adhesion, a process important for neural plasticity (Hill et al., 2012). Moreover, the knockdown of *NT5C2* implicated terms involved in heart muscle contraction and cardiomyopathy (Tables 4 and 5). Noteworthy, chromosome 10q24 is also genome-wide significant for blood pressure (Newton-Cheh et al., 2009), coronary artery disease (Schunkert et al., 2011), fat accumulation (Hotta et al., 2012) and body mass index (Wen et al., 2014). It appears that variation in *NT5C2* might be associated with pleiotropic effects. The schizophrenia risk indel rs11191419 on chromosome 10q24 is in strong LD with rs12413409 (r^2=0.82), which shows genome-wide significant association with coronary artery disease (Schunkert et al., 2011), and rs11191548 (r^2=0.82), which shows genome-wide significant association with blood pressure (Newton-Cheh et al., 2009), for which risk alleles are associated with increased susceptibility to all such conditions. The comorbidity between schizophrenia and cardiac disease is well documented (Hou et al., 2015; Jindal et al., 2005). Schizophrenia patients show a reduction in life expectancy from 10-20 years, which is associated with natural causes, but especially with cardiovascular disease (Owen et al., 2016). Patients are three times more likely to die of sudden death (Jindal et al., 2005), and show 30% higher risk of dying of coronary heart disease (Hennekens et al., 2005), as compared to a member of the general population. It is possible, however, that patients frequently suffer from cardiac conditions as they may adopt lifestyles which predispose to obesity and increased risk for hypertension, smoking, insulin resistance and diabetes (Hennekens et al., 2005). Nonetheless, genetic mutations implicated in multiple diseases are not rare, with one example being the genetic variants in *CACNA1C* (calcium channel, voltage-dependent, L type, alpha 1C subunit). Mutations in this gene were initially implicated in cardiac deficits and mental retardation in Timothy's syndrome (Liao and Soong, 2010). Recently, however, variants in *CACNA1C* have been robustly associated with risk to several

psychiatric conditions, including schizophrenia, bipolar disorder and major depression (He et al., 2014).

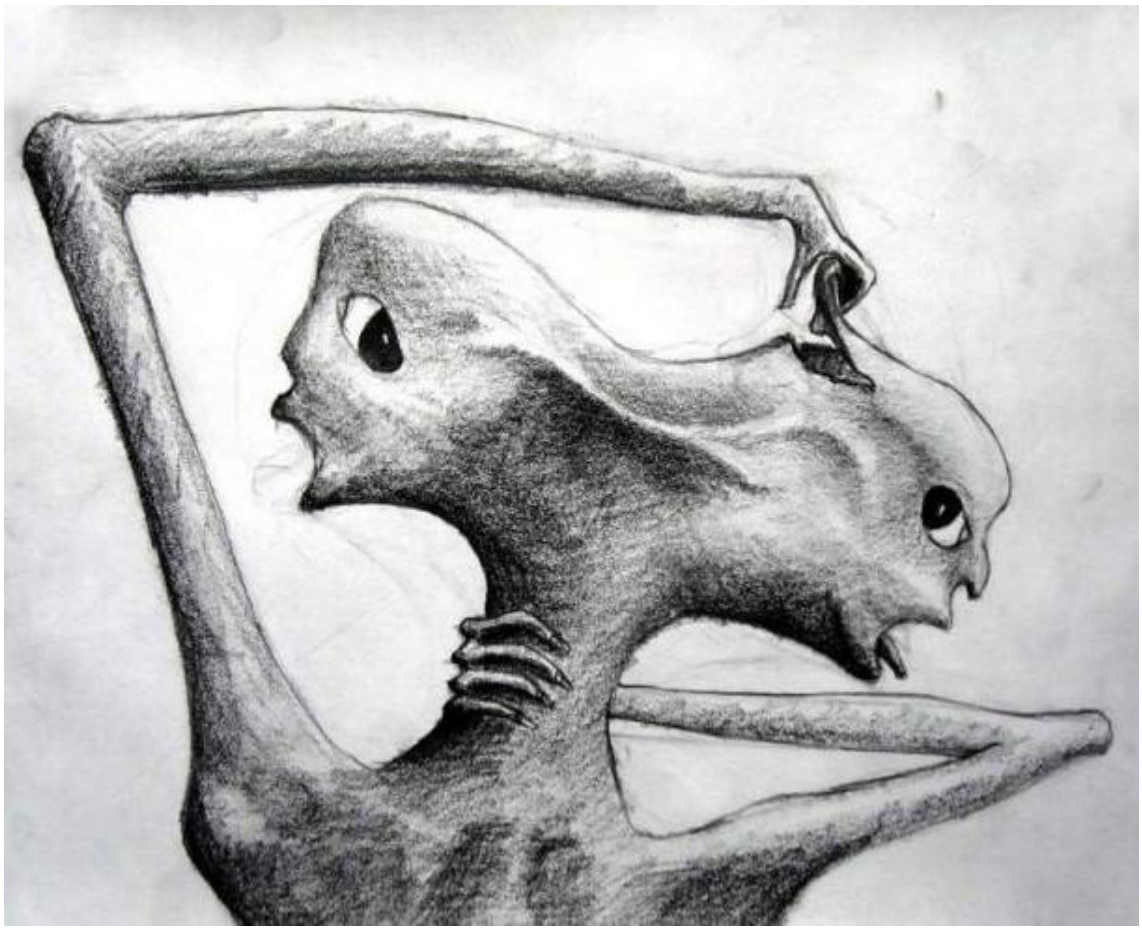
Interestingly, the genes involved in cardiac contraction significantly affected by the knockdown of *NT5C2* include *TPM1* and *TPM2*, which encode isoforms of tropomyosin. More than 40 isoforms of tropomyosin have been found, produced by four genes (*TPM1-4).* These isoforms have been suggested to participate at the cytoskeleton in several tissues, at time- and location-specific occasions (Gunning et al., 2008). Curiously, products of *TPM3* and *TPM4* have been found in the postsynaptic density region (Guven et al., 2011). Moreover, the connectivity mapping analysis predicted at least two cardiac steroids, strophanthidin and proscillaridin, inhibitors of the Na^+/K^+ ATPase (Denicolai et al., 2014; Song et al., 2014), to be associated with a similar transcriptomic signature as the one elicited by the knockdown of *NT5C2* in NPCs. Inhibition of the Na^+/K^+ ATPase by these drugs (or by hypoxia, an event relevant to schizophrenia during neurodevelopment) is expected to cause intracellular accumulation of Ca^{2+} (Gusarova et al., 2011), which in muscle cells may trigger contractions, but in nervous cells might disrupt action potentials.

Pathways related to cancer were also found over-represented in association with the knockdown of *NT5C2*. This was supported by connectivity mapping, which indicated drugs with antitumour properties such as anandamide (De Petrocellis et al., 1998) and hippeastrine (Cao et al., 2013) to counteract the putative schizophrenia risk mechanism associated with the decreased expression of *NT5C2*. The literature, however, contains contradictory findings regarding susceptibility to cancer in schizophrenia patients. In a cohort of 59,233 schizophrenia individuals in Sweden, for example, Ji et al. (2013) reported decreased incidence of cancer in patients and their unaffected relatives. In a cohort of 29,996 schizophrenia individuals in Finland, however, Lichtermann et al. (2001) found increased incidence of cancer in patients, and decreased risk in relatives. The heterogeneity intrinsic to both conditions hinders the investigation of common genetic grounds between these disorders, while this comorbidity might be additionally explained by lifestyle.

As for potential downstream mechanisms associated with the transient knockdown of *NT5C2* in the neural progenitor cells, gene ontology analysis suggested the activation of the mitogen-activated protein kinases (MAPK) signalling. MAPK pathways have been implicated in various processes that are relevant to schizophrenia, including neural plasticity, cellular proliferation, differentiation, cell death and neuronal function (Meloche and Pouyssegur, 2007; Roux and Blenis, 2004; Stephan et al., 2009; Thomas and Huganir, 2004). The members of the MAPK cascade constitute a conserved family of serine/threonine protein kinases. The kinases *RACK1, FYN* and *CDK5*, for example, were found over-represented in the *post-mortem* DLPFC of schizophrenia patients (Funk et al., 2012). Activation of MAPK signalling has also been observed increased in the cerebellar vermis and thalamus of schizophrenia patients, but not in other brain regions, suggesting that its abnormal regulation in schizophrenia might occur in specific brain areas (Kyosseva et al., 1999; Kyosseva et al., 2000; Kyosseva, 2004). The ubiquity of MAPK signalling molecules, and their ability to integrate external stimuli to different neurotransmitter systems, grants this pathway a privileged connection with such a heterogeneous condition that is schizophrenia (Funk et al., 2012). Although the mean increased phosphorylation at Thr202/Tyr204 from MAPK-ERK1/2 was observed on the knockdown, this event did not statistically differ from controls (Figure 8). It is possible that downstream effects associated with *NT5C2* knockdown might be exerted by other members of the MAPK signalling, such as the ones encoded by the other MAPK genes that were observed to be differentially expressed by the silencing. Activation of another signalling cascade, the AMP-activated protein kinases (AMPK), was also investigated in NPCs subject to *NT5C2* knockdown, as the cytosolic 5'-nucleotidase II has been hypothesised to regulate intracellular nucleotide and ribonucleotide pools (Itoh, 2013). Increased phosphorylation of AMPK-alpha at residue Thr172, indicating activation of AMPK signalling, has been previously observed to be associated with the knockdown of *NT5C2*. Interestingly, gene expression profiling of neurons derived from schizophrenia patients' hiPSCs showed altered expression of components of

cyclic AMP pathways (Brennand et al., 2011). In human and rodent skeletal muscle fibres, moreover, knockdown of *NT5C2* has been observed to cause activation of AMPK, and to trigger a series of energy mobilisation events (Kulkarni et al., 2011). AMPK has additionally been observed to activate MAPK-p38 in the ischemic muscle heart (Li et al., 2005). The role of these kinases in the downregulation of *NT5C2* in neural tissue, as a putative risk mechanism for schizophrenia remains unclear.

As for the connectivity mapping analysis, promising evidence to support its use for advancing therapeutics in other fields of medicine has been found (Lamb, 2007). Although common alleles implicated in risk for psychiatric diseases are typically associated with small effects, it is possible that these converge onto shared downstream cellular pathways, which in turn could be targeted by repositioned drugs. Interestingly, the connectivity mapping analysis suggested the typical antipsychotic fluphenazine and the atypical antipsychotic ziprasidone as the top chemical compounds predicted to counteract the *NT5C2* knockdown in NPCs (Table 6). As these antipsychotics bind to multiple targets, it is possible that some of these will trigger changes in the intracellular purine nucleotide levels. Many other drugs were additionally suggested to reverse the transcriptomic signature associated with the knockdown, and more studies are warranted to investigate their ability in doing so. Results obtained in this chapter suggest a potential role for *NT5C2* in mediating risk for schizophrenia and cardiac disease, possibly by alterations to AMPK signalling. While gene ontology analysis and connectivity mapping methods seem far from entirely capturing the complex nature of schizophrenia, they may shed light into biological processes implicated in disease and potential drugs to be used to counteract them. While these findings need further validation and replication *in vitro* and *in vivo*, it is hypothesised that the functional characterisation of several risk mechanisms for schizophrenia may expose converging molecular pathways involved in risk, which may expose novel drug targets. This is particularly promising in a field where nearly the same antipsychotic drugs prescribed in the 1950s are still used today despite severe side effects and variable efficacy (Kapur and Mamo, 2003).



Drawing from an unknown artist, supposedly found in a psychiatric institution.

Chapter 6

Discussion

6.1. Summary of findings

Genome-wide association studies have provided an unparalleled starting point for the study of common risk mechanisms associated with schizophrenia. Increasing GWAS cohorts have so far implicated 128 common variants, spread over 108 genetic loci, in risk for schizophrenia. As for most GWAS, however, implicated risk variants are mostly non-coding and therefore hypothesised to impact on gene expression by means of cis-regulatory effects, potentially elicited on rather distant targets (long-range cis-regulatory effects). The identification of the causal risk mechanisms is additionally hindered by phenomena such as linkage disequilibrium, location- and developmental stage-dependent *cis*-regulatory effects, or transcript-specific mechanisms. Their identification, however, may expose shared downstream effects associated with disease, and consequently novel drug targets for treatment.

In this context, the main objective of this thesis was to functionally characterise one of the earliest and most robust genome-wide association signals to emerge from large-scale GWAS of schizophrenia, located on chromosome 10q24 (Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Schizophrenia Working Group of the PGC, 2014). Its association with schizophrenia has been replicated elsewhere (Aberg et al., 2013; Guan et al., 2016), and variation at the locus has been suggested to contribute to risk to psychiatric conditions in general, including depression, bipolar disorder, attention-deficit hyperactivity disorder and autism (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). Risk variants at this locus have additionally been implicated in morphometrical changes in brain areas (Ohi et al., 2013; Ohi, 2015) and in indexes of social cognition (Rose et al., 2014) putatively affected in schizophrenia. The molecular mechanisms underpinning these events, however, remain unclear.

The non-coding schizophrenia risk variants on chromosome 10q24, rs11191419 and chr10_104957618_I, were found to elicit multiple *cis*-regulatory effects at this locus (Chapter 2). Such *cis*-regulatory effects were found to be operating on *BORCS7, AS3MT*, and *NT5C2*, in different brain tissues implicated in

schizophrenia: the developing foetal brain, and the adult dorsolateral prefrontal cortex, hippocampus and caudate (Duarte et al., 2016). In general terms, the risk allele of rs11191419 was associated with increased expression of *BORCS7* and *AS3MT,* and with reduced expression of *NT5C2*, while the risk allele of the indel chr10_104957618_I was found to be associated with reduced expression of all these in all brain tissues. Corroborating these findings, Li et al. (2016) found that risk genotype at rs7085104 (which is in LD with rs11191419: r^2=0.79) was associated with increased expression of *BORCS7* and *AS3MT* in several brain areas. The authors note, however, that rs7085104 is actually in LD with other variants at the locus distinctly driving expression of these genes: a variable number tandem repeat (r^2=0.94), which is regulating expression of the putative schizophrenia risk transcript of *AS3MT (AS3MT^{d2d3})*, and rs11441374 (r^2=0.70), which is driving expression of *BORCS7*. As for *NT5C2*, corroborating the results presented in Chapter 2, Hauberg et al. (2016) found that the risk allele of rs11191548 (which is in LD with the indel chr10_104957618_I: r^2=0.82) provides a binding site for miRNA-206 located in the 3' UTR of this gene, and is therefore associated with reduced *NT5C2* transcript levels. *NT5C2* was further explored in the next chapters since both risk variants on chromosome 10q24 were found to concordantly reduce its expression in multiple brain areas.

Methodological limitations in Chapter 2 precluded the identification of specific transcripts of *BORCS7, AS3MT,* or *NT5C2* that were subject to the *cis*-regulatory mechanisms associated with schizophrenia. Therefore, the second aim of this thesis was to investigate, in Chapter 3, which transcripts of these genes were expressed in the adult DLPFC and the foetal brain, which were contributing to the allelic expression mechanism detected previously. The pilot RNA-seq study suggested that RefSeq transcripts of these genes were the main variants produced in these brain tissues, although novel, less expressed transcripts also exist. These novel transcripts include the putative schizophrenia susceptibility transcript *AS3MT^{d2d3}* (Li et al., 2016), as well as the novel *NT5C2* variants, *NT5C2$^{e3.1}$*, *NT5C2^{d3}*, and *NT5C2$^{e3.1\ d3}$*. The new *NT5C2*

transcripts were experimentally confirmed in different brain areas. The RNA-seq data provided evidence to suggest that the full length transcripts of *NT5C2* and *AS3MT* were more highly abundant than the novel variants, and therefore more likely to contribute to the cis-regulatory signal observed in Chapter 2. More studies are warranted, however, to validate these findings. In any case, the sequence encoding the nucleotidase domain was not expected to be affected in the new transcripts according to analysis of protein motifs (Pfam) in the UCSC Browser. It is possible that these novel transcripts provide differences in the amino-terminus of the protein, perhaps affecting turnover or its affinity for regulatory molecules.

The role of *NT5C2* in the nervous system was unclear prior to this study. Therefore, its distribution in the adult DLPFC and in NPCs was investigated in Chapter 4. The 5'-nucleotidase cytosolic II, enzyme encoded by *NT5C2*, was found to be dispersed through the soma and cellular processes of the NPCs, and was additionally found in neurons, neuropil and glial cells in the adult DLPFC. It is possible that NT5C2 is differentially distributed according to cell type in this brain region, as a previous study suggested that this enzyme is more abundantly expressed in neurons rather than glial cells, at least in the cerebral cortex (Uhlen et al., 2015).

The final objective of this thesis, in Chapter 5, was to investigate the global gene expression changes associated with the decreased expression of *NT5C2* in NPCs, as a proxy for modelling a putative schizophrenia risk mechanism *in vitro*. The siRNA-mediated transient knockdown of *NT5C2* was confirmed by means of immuno-cytochemistry and qPCR, and therefore the effects on global gene expression were likely to be a consequence of reduced protein and transcript levels. The knockdown affected the expression of 65 genes of known function, which revealed enrichment for GO and KEGG terms related to the actin cytoskeleton, heart muscle contraction and cancer, at an FDR<0.05. Strikingly, genome-wide significant signals for coronary artery disease, rs12413409 (Dichgans et al., 2014; Schunkert et al., 2011; Takeuchi et al., 2015), and for blood pressure, rs11191548 (Newton-Cheh et al., 2009), are in strong LD with the schizophrenia risk variant rs11191419 (r^2=0.82), with major

alleles associated with increased risk for all these conditions and with reduced expression of *NT5C2,* suggesting that genetic variation in this gene is associated with pleiotropic effects. Gene ontology analysis additionally revealed an enrichment for the activation of the mitogen-activated protein kinases (MAPK) signalling, at an FDR<0.1. Given the pivotal role of MAPK signalling in neural plasticity, cellular differentiation and cell death (Meloche and Pouyssegur, 2007; Roux and Blenis, 2004; Stephan et al., 2009; Thomas and Huganir, 2004), activation of MAPK signalling in association with the knockdown was assessed by Western blotting, as function of phosphorylation of ERK1/2 (p42/p44) at the residues Thr202/Tyr204. The knockdown was associated with increased mean phosphorylation at these residues, but this event was not statistically significant – although it is possible that other members of the MAPK cascade are involved instead. Phosphorylation at Thr172 of AMPK-alpha, however, which is a member of the AMP-activated protein kinases (AMPK) signalling cascade, was significantly associated with the knockdown. Activation of AMPK signalling in human and rodent skeletal muscle fibres, for which expression of *NT5C2* had been transiently reduced, has been previously reported (Kulkarni et al., 2011). Moreover, AMPK and MAPK signalling cascades are likely to cross-talk, as AMPK signalling has been found regulate activation of p38-MAPK in the ischemic heart muscle (Li et al., 2005). Finally, connectivity mapping was applied to the gene set implicated in the knockdown, which predicted fluphenazine and ziprasidone, respectively typical and atypical antipsychotics, as top ranked drugs for reversing the transcriptional effects of NT5C2 knockdown in NPCs. Many other drugs were additionally suggested to reverse the transcriptomic signature associated with the knockdown, and more studies are warranted to investigate their ability in doing so. It is possible that other risk mechanisms for schizophrenia may affect AMPK or MAPK signalling. Therefore, it will be fundamental to validate these findings in further *in vitro* and *in vivo* studies, and to establish a phenotype associated with the knockdown to investigate the effect of the drugs predicted by connectivity mapping in reversing it.

6.2. Implications of findings

This thesis represents the transitioning of schizophrenia research to the 'post-GWAS Era'. The hypothesis-free genome-wide association studies provided a valuable starting point for the investigation of schizophrenia aetiology, by exposing common genomic loci involved in risk. The *cis*-regulatory effects associated with the schizophrenia risk alleles on chromosome 10q24 have been suggested to impact on the expression of *BORCS7, AS3MT*, and *NT5C2*, which has been replicated by other groups (Hauberg et al., 2016; Li et al., 2016), although other genes at the locus might also be affected via long-range cis-regulatory effects. The identification of downstream biological mechanisms associated with risk conferred via genetic variation altering *NT5C2* expression has the potential to improve knowledge of disease aetiology, and to expose novel drug targets. The global gene expression changes caused by decreased levels of *NT5C2* was therefore recapitulated *in vitro* using neural stem cells. Evidence for the involvement of processes associated with the cytoskeleton, cellular metabolism, cellular adhesion, and the AMPK signalling cascades were observed, but the exact mechanisms underpinning these events remain unclear. This study supports, nonetheless, the purinergic hypothesis of schizophrenia (Lara and Souza, 2000). Interestingly, allopurinol (an agonist of adenosine) has been shown to improve symptoms associated with schizophrenia (Lara et al., 2001) and cardiac disease (Kelkar et al., 2011).

It is estimated that hundreds of rare and common genetic variants play a role in schizophrenia susceptibility (Bray and Hill, 2016; Lee et al., 2012; Roussos et al., 2014; Tansey et al., 2016). The functional characterisation of a single locus of small effect obviously cannot capture the complexity of this debilitating disease. However, it is plausible that groups of risk alleles and environmental cues might converge onto hub pathways, which may be impaired in specific brain regions, in peculiar time points of development, or in subgroups of patients. There are several implications for diagnosis, prevention and treatment of schizophrenia associated with this idea, raising the necessity for the characterisation of individual risk mechanisms.

6.3. Future directions

Further experiments that could build on the work in this thesis include:

- Quantifying specific *NT5C2* transcripts in a case-control design study, to investigate the association of specific transcripts with disease. This could be performed using publicly available RNA-seq data.

- Determining a quantifiable cellular phenotype associated with the knockdown of *NT5C2* in NPCs, e.g. via markers of proliferation or neuronalisation by means of high throughput image screening (e.g. using the Cell Insight®, ThermoFisher Scientific).

- Investigating the ability of fluphenazine, ziprasidone and other drugs predicted to reverse the transcriptional effects of *NT5C2* knockdown in reversing the physiological phenotype associated with *NT5C2* silencing.

- Further exploring the role of AMPK signalling in mediating risk for schizophrenia, as well as its role in maintenance or regulation of the cytoskeleton and metabolic processes.

6.4. Concluding remarks

Schizophrenia is a complex disorder which is associated with hundreds of common and rare genetic risk variants which act in concert with environmental factors to dictate risk to disease. By describing individual risk mechanisms, it may be possible to improve diagnosis, treatment or prevention options for susceptible or affected individuals in the future. In this thesis, schizophrenia risk alleles on chromosome 10q24 were found to be associated with *cis*-regulatory effects on *BORCS7, AS3MT* and *NT5C2* in the human brain. All RefSeq transcripts of these genes, as well as novel variants of *AS3MT* (*AS3MT*^{d2d3}) and *NT5C2* (*NT5C2*^{d3}, *NT5C2*$^{e3.1}$, *NT5C2*$^{e3.1\ d3}$) were detected by RNA-seq in the adult DLPFC and in the foetal brain. RefSeq and novel transcript variants of *NT5C2* were experimentally validated in several other brain samples. The expression of the previously annotated transcripts was higher than the novel variants, suggesting that they provide a larger contribution towards the allelic expression signal associated with risk observed in Chapter 2. The localisation of the enzyme produced by *NT5C2,* the 5'-cytosolic nucleotidase II, was investigated in the adult DLPFC and in neural progenitor cells. NT5C2 was found to be dispersed through the soma and cellular processes of NPCs, and it was also observed in all neural cell types in the adult DLPFC. Finally, the risk mechanism associated with the downregulation of *NT5C2* was reproduced *in vitro* with the use of small interfering RNAs in cultures of human neural progenitor cells. The regulation of biological pathways related to cytoskeleton, cellular adhesion and metabolic processes, including differential activation of AMPK signalling, were suggested to be altered by the knockdown of *NT5C2*. Further studies are warranted to confirm these findings, which may improve our understanding of the aetiology of schizophrenia and might expose novel drug targets.

References

1000 Genomes Project Consortium, 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491(7422), 56-65.

Aberg, K.A., Liu, Y., Bukszar, J., McClay, J.L., Khachane, A.N., Andreassen, O.A., et al., 2013. A comprehensive family-based replication study of schizophrenia genes. JAMA Psychiatry 70(6), 573-581.

Ackenheil, M., Weber, K., 2004. Differing response to antipsychotic therapy in schizophrenia: pharmacogenomic aspects. Dialogues Clin Neurosci 6(1), 71-77.

Adriano, F., Caltagirone, C., Spalletta, G., 2012. Hippocampal volume reduction in first-episode and chronic schizophrenia: a review and meta-analysis. Neuroscientist 18(2), 180-200.

American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders: DSM-V (5th edition), Washington, DC. APA.

Andersen, C.L., Jensen, J.L., Orntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64(15), 5245-5250.

Anderson, G.W., Deans, P.J., Taylor, R.D., Raval, P., Chen, D., Lowder, H., et al., 2015. Characterisation of neurons derived from a cortical human neural stem cell line CTX0E16. Stem Cell Res Ther 6, 149.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arjona, F.J., de Baaij, J.H., Schlingmann, K.P., Lameris, A.L., van Wijk, E., Flik, G., et al., 2014. CNNM2 mutations cause impaired brain development and seizures in patients with hypomagnesemia. PLoS Genet 10(4), e1004267.

Arnold, S.J., Ivleva, E.I., Gopal, T.A., Reddy, A.P., Jeon-Slaughter, H., Sacco, C.B., et al., 2015. Hippocampal volume is reduced in schizophrenia and schizoaffective disorder but not in psychotic bipolar I disorder demonstrated by both manual tracing and automated parcellation (FreeSurfer). Schizophr Bull 41(1), 233-249.

Arnsten, A.F.T., 2013. The Neurobiology of Thought: The Groundbreaking Discoveries of Patricia Goldman-Rakic 1937–2003. Cerebral Cortex 23(10), 2269-2281.

Baker, M., 2015. Reproducibility crisis: Blame it on the antibodies. Nature 521(7552), 274-276.

Bakhshi, K., Chance, S.A., 2015. The neuropathology of schizophrenia: A selective review of past studies and emerging themes in brain structure and cytoarchitecture. Neuroscience 303, 82-102.

Barr, C.L., Misener, V.L., 2016. Decoding the non-coding genome: elucidating genetic risk outside the coding genome. Genes Brain Behav 15(1), 187-204.

Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2), 263-265.

Bassett, A.S., Chow, E.W.C., 2008. Schizophrenia and 22q11.2 Deletion Syndrome. Current psychiatry reports 10(2), 148-157.

Belforte, J.E., Zsiros, V., Sklar, E.R., Jiang, Z., Yu, G., Li, Y., et al., 2010. Postnatal NMDA receptor ablation in corticolimbic interneurons confers schizophrenia-like phenotypes. Nat Neurosci 13(1), 76-83.

Berger, A., 1999. What are leukotrienes and how do they work in asthma? BMJ 319(7202), 90.

Berk, M., Copolov, D., Dean, O., Lu, K., Jeavons, S., Schapkaitz, I., et al., 2008. N-acetyl cysteine as a glutathione precursor for schizophrenia--a double-blind, randomized, placebo-controlled trial. Biol Psychiatry 64(5), 361-368.

Bianchi, V., Spychala, J., 2003. Mammalian 5'-nucleotidases. J Biol Chem 278(47), 46195-46198.

Bird, J.M., 1985. Computed tomographic brain studies and treatment response in schizophrenia. Can J Psychiatry 30(4), 251-254.

Birnbaum, R., Jaffe, A.E., Chen, Q., Hyde, T.M., Kleinman, J.E., Weinberger, D.R., 2015. Investigation of the prenatal expression patterns of 108 schizophrenia-associated genetic loci. Biol Psychiatry 77(11), e43-51.

Bitanihirwe, B.K.Y., Woo, T.-U.W., 2011. Oxidative Stress in Schizophrenia: An Integrated Approach. Neuroscience and biobehavioral reviews 35(3), 878-893.

Bleich, A., Brown, S.L., Kahn, R., van Praag, H.M., 1988. The role of serotonin in schizophrenia. Schizophr Bull 14(2), 297-315.

Blomström, Å., Karlsson, H., Gardner, R., Jörgensen, L., Magnusson, C., Dalman, C., 2016. Associations Between Maternal Infection During Pregnancy, Childhood Infections, and the Risk of Subsequent Psychotic Disorder—A Swedish Cohort Study of Nearly 2 Million Individuals. Schizophrenia Bulletin 42(1), 125-133.

Boga Pekmezekmek, A., Binokay, U.S., Secilmis, M.A., Kumcu, E., Simsek, E., Akillioglu, K., et al., 2015. Evaluating the Teratogenicity of Ritodrine and Nifedipine using a Frog Embryo Teratogenesis assay (FETAX). Drug Chem Toxicol 38(3), 254-265.

Boison, D., Singer, P., Shen, H.Y., Feldon, J., Yee, B.K., 2012. Adenosine hypothesis of schizophrenia--opportunities for pharmacotherapy. Neuropharmacology 62(3), 1527-1543.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15), 2114-2120.

Bray, N.J., Owen, M.J., 2001. Searching for schizophrenia genes. Trends Mol Med 7(4), 169-174.

Bray, N.J., Buckland, P.R., Williams, N.M., Williams, H.J., Norton, N., Owen, M.J., et al., 2003a. A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. Am J Hum Genet 73(1), 152-161.

Bray, N.J., Buckland, P.R., Owen, M.J., O'Donovan, M.C., 2003b. Cis-acting variation in the expression of a high proportion of genes in human brain. Hum Genet 113(2), 149-153.

Bray, N.J., Preece, A., Williams, N.M., Moskvina, V., Buckland, P.R., Owen, M.J., et al., 2005. Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression. Hum Mol Genet 14(14), 1947-1954.

Bray, N.J., 2008. Gene expression in the etiology of schizophrenia. Schizophr Bull 34(3), 412-418.

Bray, N.J., Leweke, F.M., Kapur, S., Meyer-Lindenberg, A., 2010. The neurobiology of schizophrenia: new leads and avenues for treatment. Curr Opin Neurobiol 20(6), 810-815.

Bray, N.J., Hill, M.J., 2016. Translating Genetic Risk Loci Into Molecular Risk Mechanisms for Schizophrenia. Schizophrenia Bulletin 42(1), 5-8.

Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., et al., 2011. Modelling schizophrenia using human induced pluripotent stem cells. Nature 473(7346), 221-225.

Buonocore, F., Hill, M.J., Campbell, C.D., Oladimeji, P.B., Jeffries, A.R., Troakes, C., et al., 2010. Effects of cis-regulatory variation differ across regions of the adult human brain. Hum Mol Genet 19(22), 4490-4496.

Callicott, J.H., Bertolino, A., Mattay, V.S., Langheim, F.J., Duyn, J., Coppola, R., et al., 2000. Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia revisited. Cereb Cortex 10(11), 1078-1092.

Cao, H., Dixson, L., Meyer-Lindenberg, A., Tost, H., 2016. Functional connectivity measures as schizophrenia intermediate phenotypes: advances, limitations, and future directions. Current Opinion in Neurobiology 36, 7-14.

Cao, Z., Yang, P., Zhou, Q., 2013. Multiple biological functions and pharmacological effects of lycorine. Science China Chemistry 56(10), 1382-1391.

Careddu, M.G., Allegrini, S., Pesi, R., Camici, M., Garcia-Gil, M., Tozzi, M.G., 2008. Knockdown of cytosolic 5'-nucleotidase II (cN-II) reveals that its activity is essential for survival in astrocytoma cells. Biochim Biophys Acta 1783(8), 1529-1535.

Carlsson, A., 1978. Antipsychotic drugs, neurotransmitters, and schizophrenia. Am J Psychiatry 135(2), 165-173.

Cazzullo, C.L., Scarone, S., Grassi, B., Vismara, C., Trabattoni, D., Clerici, M., et al., 1998. Cytokines production in chronic schizophrenia patients with or without paranoid behaviour. Prog Neuropsychopharmacol Biol Psychiatry 22(6), 947-957.

Chen, X., Weaver, J., Bove, B.A., Vanderveer, L.A., Weil, S.C., Miron, A., et al., 2008. Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. Hum Mol Genet 17(9), 1336-1348.

Cividini, F., Filoni, D.N., Pesi, R., Allegrini, S., Camici, M., Tozzi, M.G., 2015a. IMP–GMP specific cytosolic 5’-nucleotidase regulates nucleotide pool and prodrug metabolism. Biochimica et Biophysica Acta (BBA) - General Subjects 1850(7), 1354-1361.

Cividini, F., Cros-Perrial, E., Pesi, R., Machon, C., Allegrini, S., Camici, M., et al., 2015b. Cell proliferation and drug sensitivity of human glioblastoma cells are altered by the stable modulation of cytosolic 5'-nucleotidase II. Int J Biochem Cell Biol 65, 222-229.

Cocks, G., Curran, S., Gami, P., Uwanogho, D., Jeffries, A.R., Kathuria, A., et al., 2014. The utility of patient specific induced pluripotent stem cells for the modelling of Autistic Spectrum Disorders. Psychopharmacology (Berl) 231(6), 1079-1088.

Cooper, B., 2005. Immigration and schizophrenia: the social causation hypothesis revisited. The British Journal of Psychiatry 186(5), 361-363.

Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 381(9875), 1371-1379.

Curley, A.A., Arion, D., Volk, D.W., Asafu-Adjei, J.K., Sampson, A.R., Fish, K.N., et al., 2011. Cortical deficits of glutamic acid decarboxylase 67 expression in schizophrenia: clinical, protein, and cell type-specific features. Am J Psychiatry 168(9), 921-929.

Davis, K.L., Nappi, J.M., 2003. The cardiovascular effects of eplerenone, a selective aldosterone-receptor antagonist. Clin Ther 25(11), 2647-2668.

De Petrocellis, L., Melck, D., Palmisano, A., Bisogno, T., Laezza, C., Bifulco, M., et al., 1998. The endogenous cannabinoid anandamide inhibits human breast cancer cell proliferation. Proc Natl Acad Sci U S A 95(14), 8375-8380.

Denicolai, E., Baeza-Kallee, N., Tchoghandjian, A., Carre, M., Colin, C., Jiglaire, C.J., et al., 2014. Proscillaridin A is cytotoxic for glioblastoma cell lines and controls tumor xenograft growth in vivo. Oncotarget 5(21), 10934-10948.

Dichgans, M., Malik, R., Konig, I.R., Rosand, J., Clarke, R., Gretarsdottir, S., et al., 2014. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. Stroke 45(1), 24-36.

Dima, D., Breen, G., 2015. Polygenic risk scores in imaging genetics: Usefulness and applications. J Psychopharmacol 29(8), 867-871.

Du, P., Kibbe, W.A., Lin, S.M., 2008. lumi: a pipeline for processing Illumina microarray. Bioinformatics 24(13), 1547-1548.

Duarte, R.R.R., Troakes, C., Nolan, M., Srivastava, D.P., Murray, R.M., Bray, N.J., 2016. Genome-wide significant schizophrenia risk variation on chromosome 10q24 is associated with altered cis-regulation of BORCS7, AS3MT, and NT5C2 in the human brain. Am J Med Genet B Neuropsychiatr Genet 171B, 806–814.

Dudbridge, F., 2013. Power and predictive accuracy of polygenic risk scores. PLoS Genet 9(3), e1003348.

Evan, G.I., Lewis, G.K., Ramsay, G., Bishop, J.M., 1985. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. Mol Cell Biol 5(12), 3610-3616.

Feinberg, I., 1982. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? J Psychiatr Res 17(4), 319-334.

Fillman, S.G., Cloonan, N., Catts, V.S., Miller, L.C., Wong, J., McCrossin, T., et al., 2013. Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. Mol Psychiatry 18(2), 206-214.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., et al., 2014. De novo mutations in schizophrenia implicate synaptic networks. Nature 506(7487), 179-184.

Funk, A.J., McCullumsmith, R.E., Haroutunian, V., Meador-Woodruff, J.H., 2012. Abnormal activity of the MAPK- and cAMP-associated signaling pathways in frontal cortical areas in postmortem brain in schizophrenia. Neuropsychopharmacology 37(4), 896-905.

Galindo, L., Pastoriza, F., Bergé, D., Mané, A., Roé, N., Pujol, N., et al., 2016. Abnormal connectivity in dorsolateral prefrontal cortex in schizophrenia patients and unaffected relatives. European Psychiatry 33, S100-S101.

Georgieva, L., Rees, E., Moran, J.L., Chambert, K.D., Milanova, V., Craddock, N., et al., 2014. De novo CNVs in bipolar affective disorder and schizophrenia. Hum Mol Genet 23(24), 6677-6683.

Ghiani, C.A., Starcevic, M., Rodriguez-Fernandez, I.A., Nazarian, R., Cheli, V.T., Chan, L.N., et al., 2010. The dysbindin-containing complex (BLOC-1) in brain: developmental regulation, interaction with SNARE proteins and role in neurite outgrowth. Mol Psychiatry 15(2), 115, 204-115.

Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.L., et al., 2010. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 6(5), e1000952.

Gilda, J.E., Ghosh, R., Cheah, J.X., West, T.M., Bodine, S.C., Gomes, A.V., 2015. Western Blotting Inaccuracies with Unverified Antibodies: Need for a Western Blotting Minimal Reporting Standard (WBMRS). PLoS One 10(8), e0135392.

Glantz, L.A., Lewis, D.A., 2000. Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. Arch Gen Psychiatry 57(1), 65-73.

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17(6), 333-351.

Gottesman, I., 1991. Schizophrenia Genesis: The Origins of Madness. Freeman, San Francisco.

Grayson, D.R., Jia, X., Chen, Y., Sharma, R.P., Mitchell, C.P., Guidotti, A., et al., 2005. Reelin promoter hypermethylation in schizophrenia. Proc Natl Acad Sci U S A 102(26), 9341-9346.

GTEx Consortium, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348(6235), 648-660.

Guan, F., Zhang, T., Li, L., Fu, D., Lin, H., Chen, G., et al., 2016. Two-stage replication of previous genome-wide association studies of AS3MT-CNNM2-NT5C2 gene cluster region in a large schizophrenia case-control sample from Han Chinese population. Schizophr Res.

Gunning, P., O'Neill, G., Hardeman, E., 2008. Tropomyosin-based regulation of the actin cytoskeleton in time and space. Physiol Rev 88(1), 1-35.

Gur, R.E., Gur, R.C., 2010. Functional magnetic resonance imaging in schizophrenia. Dialogues Clin Neurosci 12(3), 333-343.

Gusarova, G.A., Trejo, H.E., Dada, L.A., Briva, A., Welch, L.C., Hamanaka, R.B., et al., 2011. Hypoxia leads to Na,K-ATPase downregulation via Ca(2+) release-activated Ca(2+) channels and AMPK activation. Mol Cell Biol 31(17), 3546-3556.

Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., et al., 1983. A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306(5940), 234-238.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., et al., 2016. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 48(3), 245-252.

Guven, K., Gunning, P., Fath, T., 2011. TPM3 and TPM4 gene products segregate to the postsynaptic region of central nervous system synapses. Bioarchitecture 1(6), 284-289.

Hakak, Y., Walker, J.R., Li, C., Wong, W.H., Davis, K.L., Buxbaum, J.D., et al., 2001. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proc Natl Acad Sci U S A 98(8), 4746-4751.

Hall, J., Trent, S., Thomas, K.L., O'Donovan, M.C., Owen, M.J., 2015. Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. Biol Psychiatry 77(1), 52-58.

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T.M., et al., 2016. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci 19(1), 48-54.

Harrison, P.J., Freemantle, N., Geddes, J.R., 2003. Meta-analysis of brain weight in schizophrenia. Schizophr Res 64(1), 25-34.

Harrison, P.J., Weinberger, D.R., 2005. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Mol Psychiatry 10(1), 40-68; image 45.

Hasan, A., Falkai, P., Wobrock, T., Lieberman, J., Glenthoj, B., Gattaz, W.F., et al., 2012. World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Schizophrenia, part 1: update 2012 on the acute treatment of schizophrenia and the management of treatment resistance. World J Biol Psychiatry 13(5), 318-378.

Hashimoto, K., 2010. Glycine transport inhibitors for the treatment of schizophrenia. Open Med Chem J 4, 10-19.

Hauberg, M.E., Holm-Nielsen, M.H., Mattheisen, M., Askou, A.L., Grove, J., Borglum, A.D., et al., 2016. Schizophrenia risk variants affecting microRNA

function and site-specific regulation of NT5C2 by miR-206. Eur Neuropsychopharmacol.

He, K., An, Z., Wang, Q., Li, T., Li, Z., Chen, J., et al., 2014. CACNA1C, schizophrenia and major depressive disorder in the Han Chinese population. Br J Psychiatry 204(1), 36-39.

Heap, G.A., Yang, J.H., Downes, K., Healy, B.C., Hunt, K.A., Bockett, N., et al., 2010. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. Hum Mol Genet 19(1), 122-134.

Hennekens, C.H., Hennekens, A.R., Hollar, D., Casey, D.E., 2005. Schizophrenia and increased risks of cardiovascular disease. Am Heart J 150(6), 1115-1121.

Hill, M.J., Bray, N.J., 2012. Evidence that schizophrenia risk variation in the ZNF804A gene exerts its effects during fetal brain development. Am J Psychiatry 169(12), 1301-1308.

Hill, M.J., Jeffries, A.R., Dobson, R.J., Price, J., Bray, N.J., 2012. Knockdown of the psychosis susceptibility gene ZNF804A alters expression of genes involved in cell adhesion. Hum Mol Genet 21(5), 1018-1024.

Hirota, T., Kishi, T., 2013. Adenosine hypothesis in schizophrenia and bipolar disorder: a systematic review and meta-analysis of randomized controlled trial of adjuvant purinergic modulators. Schizophr Res 149(1-3), 88-95.

Honea, R., Crow, T.J., Passingham, D., Mackay, C.E., 2005. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. Am J Psychiatry 162(12), 2233-2245.

Hotta, K., Kitamoto, A., Kitamoto, T., Mizusawa, S., Teranishi, H., Matsuo, T., et al., 2012. Genetic variations in the CYP17A1 and NT5C2 genes are associated with a reduction in visceral and subcutaneous fat areas in Japanese women. J Hum Genet 57(1), 46-51.

Hou, P.Y., Hung, G.C., Jhong, J.R., Tsai, S.Y., Chen, C.C., Kuo, C.J., 2015. Risk factors for sudden cardiac death among patients with schizophrenia. Schizophr Res 168(1-2), 395-401.

Huang, H.S., Akbarian, S., 2007. GAD1 mRNA expression and DNA methylation in prefrontal cortex of subjects with schizophrenia. PLoS One 2(8), e809.

Hultman, C.M., Sparén, P., Takei, N., Murray, R.M., Cnattingius, S., 1999. Prenatal and perinatal risk factors for schizophrenia, affective psychosis, and reactive psychosis of early onset: case-control study. BMJ : British Medical Journal 318(7181), 421-426.

Hunsucker, S.A., Mitchell, B.S., Spychala, J., 2005. The 5'-nucleotidases as regulators of nucleotide and drug metabolism. Pharmacol Ther 107(1), 1-30.

Insel, T.R., 2009. Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. Arch Gen Psychiatry 66(2), 128-133.

International HapMap Consortium, 2003. The International HapMap Project. Nature 426(6968), 789-796.

International Schizophrenia Consortium, Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., et al., 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460(7256), 748-752.

Itoh, R., 2013. Enzymatic properties and physiological roles of cytosolic 5'-nucleotidase II. Curr Med Chem 20(34), 4260-4284.

Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J., Kimmel, M., 2015. Microarray experiments and factors which affect their reliability. Biol Direct 10, 46.

Janda, E., Lascala, A., Carresi, C., Parafati, M., Aprigliano, S., Russo, V., et al., 2015. Parkinsonian toxin-induced oxidative stress inhibits basal autophagy in astrocytes via NQO2/quinone oxidoreductase 2: Implications for neuroprotection. Autophagy 11(7), 1063-1080.

Ji, J., Sundquist, K., Ning, Y., Kendler, K.S., Sundquist, J., Chen, X., 2013. Incidence of cancer in patients with schizophrenia and their first-degree relatives: a population-based study in Sweden. Schizophr Bull 39(3), 527-536.

Jindal, R., MacKenzie, E.M., Baker, G.B., Yeragani, V.K., 2005. Cardiac risk and schizophrenia. J Psychiatry Neurosci 30(6), 393-395.

Johnstone, E.C., Crow, T.J., Frith, C.D., Husband, J., Kreel, L., 1976. Cerebral ventricular size and cognitive impairment in chronic schizophrenia. Lancet 2(7992), 924-926.

Johnstone, R.W., 2002. Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. Nat Rev Drug Discov 1(4), 287-299.

Joyce, E.M., Roiser, J.P., 2007. Cognitive heterogeneity in schizophrenia. Curr Opin Psychiatry 20(3), 268-272.

Jung, H., Yoon, B.C., Holt, C.E., 2012. Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. Nat Rev Neurosci 13(5), 308-324.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., et al., 2011. Spatio-temporal transcriptome of the human brain. Nature 478(7370), 483-489.

Kapur, S., Mamo, D., 2003. Half a century of antipsychotics and still a central role for dopamine D2 receptors. Prog Neuropsychopharmacol Biol Psychiatry 27(7), 1081-1090.

Kelkar, A., Kuo, A., Frishman, W.H., 2011. Allopurinol as a cardiovascular drug. Cardiol Rev 19(6), 265-271.

Khandaker, G.M., Zimbron, J., Lewis, G., Jones, P.B., 2013. Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. Psychol Med 43(2), 239-257.

Khashan, A.S., Abel, K.M., McNamee, R., Pedersen, M.G., Webb, R.T., Baker, P.N., et al., 2008. Higher risk of offspring schizophrenia following antenatal maternal exposure to severe adverse life events. Arch Gen Psychiatry 65(2), 146-152.

Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., et al., 2012. Exome sequencing and the genetic basis of complex traits. Nat Genet 44(6), 623-630.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14(4), R36.

Kirkbride, J.B., Susser, E., Kundakovic, M., Kresovich, J.K., Smith, G.D., Relton, C.L., 2012. Prenatal nutrition, epigenetics and schizophrenia risk: can we test causal effects? Epigenomics 4(3), 303-315.

Konopaske, G.T., Lange, N., Coyle, J.T., Benes, F.M., 2014. PRefrontal cortical dendritic spine pathology in schizophrenia and bipolar disorder. JAMA Psychiatry 71(12), 1323-1331.

Kooyman, I., Dean, K., Harvey, S., Walsh, E., 2007. Outcomes of public concern in schizophrenia. Br J Psychiatry Suppl 50, s29-36.

Krabbendam, L., van Os, J., 2005. Schizophrenia and Urbanicity: A Major Environmental Influence—Conditional on Genetic Risk. Schizophrenia Bulletin 31(4), 795-799.

Kriaucionis, S., Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324(5929), 929-930.

Krystal, J.H., Karper, L.P., Seibyl, J.P., Freeman, G.K., Delaney, R., Bremner, J.D., et al., 1994. Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. Arch Gen Psychiatry 51(3), 199-214.

Kulkarni, S.S., Karlsson, H.K., Szekeres, F., Chibalin, A.V., Krook, A., Zierath, J.R., 2011. Suppression of 5'-nucleotidase enzymes promotes AMP-activated protein kinase (AMPK) phosphorylation and metabolism in human and mouse skeletal muscle. J Biol Chem 286(40), 34567-34574.

Kyosseva, S.V., Elbein, A.D., Griffin, W.S., Mrak, R.E., Lyon, M., Karson, C.N., 1999. Mitogen-activated protein kinases in schizophrenia. Biol Psychiatry 46(5), 689-696.

Kyosseva, S.V., Elbein, A.D., Hutton, T.L., Griffin, S.T., Mrak, R.E., Sturner, W.Q., et al., 2000. Increased levels of transcription factors Elk-1, cyclic adenosine monophosphate response element-binding protein, and activating transcription factor 2 in the cerebellar vermis of schizophrenic patients. Arch Gen Psychiatry 57(7), 685-691.

Kyosseva, S.V., 2004. Differential expression of mitogen-activated protein kinases and immediate early genes fos and jun in thalamus in schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry 28(6), 997-1006.

Lahti, A.C., Weiler, M.A., Tamara Michaelidis, B.A., Parwani, A., Tamminga, C.A., 2001. Effects of ketamine in normal and schizophrenic volunteers. Neuropsychopharmacology 25(4), 455-467.

Lamb, J., 2007. The Connectivity Map: a new tool for biomedical research. Nat Rev Cancer 7(1), 54-60.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4), 357-359.

Lara, D.R., Souza, D.O., 2000. Schizophrenia: a purinergic hypothesis. Med Hypotheses 54(2), 157-166.

Lara, D.R., Brunstein, M.G., Ghisolfi, E.S., Lobato, M.I., Belmonte-de-Abreu, P., Souza, D.O., 2001. Allopurinol augmentation for poorly responsive schizophrenia. Int Clin Psychopharmacol 16(4), 235-237.

Laruelle, M., 2000. The role of endogenous sensitization in the pathophysiology of schizophrenia: implications from recent brain imaging studies. Brain Res Brain Res Rev 31(2-3), 371-384.

Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., et al., 2016. Third-generation sequencing and the future of genomics. bioRxiv.

Lee, K., Yoshida, T., Kubicki, M., Bouix, S., Westin, C.F., Kindlmann, G., et al., 2009. Increased diffusivity in superior temporal gyrus in patients with schizophrenia: a Diffusion Tensor Imaging study. Schizophr Res 108(1-3), 33-40.

Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Schizophrenia Psychiatric Genome-Wide Association Study, C., International Schizophrenia, C., et al., 2012. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 44(3), 247-250.

Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., et al., 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45(9), 984-994.

Leucht, S., Corves, C., Arbter, D., Engel, R.R., Li, C., Davis, J.M., 2009. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. Lancet 373(9657), 31-41.

Leucht, S., Tardy, M., Komossa, K., Heres, S., Kissling, W., Salanti, G., et al., 2012. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. Lancet 379(9831), 2063-2071.

Levenson, J.M., Roth, T.L., Lubin, F.D., Miller, C.A., Huang, I.C., Desai, P., et al., 2006. Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. J Biol Chem 281(23), 15763-15773.

Lewis, D.A., Gonzalez-Burgos, G., 2008. Neuroplasticity of neocortical circuits in schizophrenia. Neuropsychopharmacology 33(1), 141-165.

Li, J., Miller, E.J., Ninomiya-Tsuji, J., Russell, R.R., 3rd, Young, L.H., 2005. AMP-activated protein kinase activates p38 mitogen-activated protein kinase by increasing recruitment of p38 MAPK to TAB1 in the ischemic heart. Circ Res 97(9), 872-879.

Li, M., Jaffe, A.E., Straub, R.E., Tao, R., Shin, J.H., Wang, Y., et al., 2016. A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. Nat Med 22, 649–656.

Liao, P., Soong, T.W., 2010. CaV1.2 channelopathies: from arrhythmias to autism, bipolar disorder, and immunodeficiency. Pflugers Arch 460(2), 353-359.

Lichtenstein, P., Yip, B.H., Bjork, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F., et al., 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 373(9659), 234-239.

Lichtermann, D., Ekelund, J., Pukkala, E., Tanskanen, A., Lonnqvist, J., 2001. Incidence of cancer among persons with schizophrenia and their relatives. Arch Gen Psychiatry 58(6), 573-578.

Lips, E.S., Cornelisse, L.N., Toonen, R.F., Min, J.L., Hultman, C.M., International Schizophrenia, C., et al., 2012. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol Psychiatry 17(10), 996-1006.

Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H., et al., 2003. Allelic variation in gene expression is common in the human genome. Genome Res 13(8), 1855-1862.

Luisada, P.V., 1978. The phencyclidine psychosis: phenomenology and treatment. NIDA Res Monogr(21), 241-253.

Maes, M., Bosmans, E., Calabrese, J., Smith, R., Meltzer, H.Y., 1995. Interleukin-2 and interleukin-6 in schizophrenia and mania: effects of neuroleptics and mood stabilizers. J Psychiatr Res 29(2), 141-152.

Mangalore, R., Knapp, M., 2007. Cost of schizophrenia in England. J Ment Health Policy Econ 10(1), 23-41.

Marino, M.J., Knutsen, L.J., Williams, M., 2008. Emerging opportunities for antipsychotic drug discovery in the postgenomic era. J Med Chem 51(5), 1077-1107.

Marwaha, S., Johnson, S., 2004. Schizophrenia and employment - a review. Soc Psychiatry Psychiatr Epidemiol 39(5), 337-349.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A., et al., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9(5), 356-369.

McGrath, J., Saha, S., Chant, D., Welham, J., 2008. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol Rev 30, 67-76.

Meloche, S., Pouyssegur, J., 2007. The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. Oncogene 26(22), 3227-3239.

Meyer, J.A., Wang, J., Hogan, L.E., Yang, J.J., Dandekar, S., Patel, J.P., et al., 2013. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nat Genet 45(3), 290-294.

Milicchio, F., Rose, R., Bian, J., Min, J., Prosperi, M., 2016. Visual programming for next-generation sequencing data analytics. BioData Min 9, 16.

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., et al., 2008. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. Am J Hum Genet 82(3), 696-711.

Millar, J.K., Wilson-Annan, J.C., Anderson, S., Christie, S., Taylor, M.S., Semple, C.A., et al., 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 9(9), 1415-1423.

Miller, C.A., Sweatt, J.D., 2007. Covalent modification of DNA regulates memory formation. Neuron 53(6), 857-869.

Mirnics, K., Middleton, F.A., Marquez, A., Lewis, D.A., Levitt, P., 2000. Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. Neuron 28(1), 53-67.

Mirnics, K., Levitt, P., Lewis, D.A., 2006. Critical appraisal of DNA microarrays in psychiatric genomics. Biol Psychiatry 60(2), 163-176.

Mita, T., Hanada, S., Nishino, N., Kuno, T., Nakai, H., Yamadori, T., et al., 1986. Decreased serotonin S2 and increased dopamine D2 receptors in chronic schizophrenics. Biol Psychiatry 21(14), 1407-1414.

Mitelman, S.A., Canfield, E.L., Newmark, R.E., Brickman, A.M., Torosjan, Y., Chu, K.-W., et al., 2009. Longitudinal Assessment of Gray and White Matter in Chronic Schizophrenia: A Combined Diffusion-Tensor and Structural Magnetic Resonance Imaging Study. The Open Neuroimaging Journal 3, 31-47.

Moller, H.J., Czobor, P., 2015. Pharmacological treatment of negative symptoms in schizophrenia. Eur Arch Psychiatry Clin Neurosci 265(7), 567-578.

Moncrieff, J., Middleton, H., 2015. Schizophrenia: a critical psychiatry perspective. Curr Opin Psychiatry 28(3), 264-268.

Monji, A., Kato, T.A., Mizoguchi, Y., Horikawa, H., Seki, Y., Kasai, M., et al., 2013. Neuroinflammation in schizophrenia especially focused on the role of microglia. Prog Neuropsychopharmacol Biol Psychiatry 42, 115-121.

Murphy, K.C., Jones, L.A., Owen, M.J., 1999. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry 56(10), 940-945.

Murray, R.M., Lewis, S.W., 1987. Is schizophrenia a neurodevelopmental disorder? Br Med J (Clin Res Ed) 295(6600), 681-682.

Myers, S., Freeman, C., Auton, A., Donnelly, P., McVean, G., 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genet 40(9), 1124-1129.

Nakazawa, K., Zsiros, V., Jiang, Z., Nakao, K., Kolata, S., Zhang, S., et al., 2012. GABAergic interneuron origin of schizophrenia pathophysiology. Neuropharmacology 62(3), 1574-1583.

Nazeri, A., Chakravarty, M.M., Felsky, D., Lobaugh, N.J., Rajji, T.K., Mulsant, B.H., et al., 2013. Alterations of superficial white matter in schizophrenia and relationship to cognitive performance. Neuropsychopharmacology 38(10), 1954-1962.

Nelson, M.D., Saykin, A.J., Flashman, L.A., Riordan, H.J., 1998. Hippocampal volume reduction in schizophrenia as assessed by magnetic resonance imaging: a meta-analytic study. Arch Gen Psychiatry 55(5), 433-440.

Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium, 2015. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci 18(2), 199-209.

Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., et al., 2009. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 41(6), 666-676.

Nichols, D.E., 2004. Hallucinogens. Pharmacol Ther 101(2), 131-181.

O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., et al., 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 40(9), 1053-1055.

Oertel-Knochel, V., Lancaster, T.M., Knochel, C., Stablein, M., Storchak, H., Reinke, B., et al., 2015. Schizophrenia risk variants modulate white matter volume across the psychosis spectrum: evidence from two independent cohorts. Neuroimage Clin 7, 764-770.

Ohi, K., Hashimoto, R., Yamamori, H., Yasuda, Y., Fujimoto, M., Umeda-Yano, S., et al., 2013. The impact of the genome-wide supported variant in the cyclin M2 gene on gray matter morphology in schizophrenia. Behav Brain Funct 9, 40.

Ohi, K., 2015. Influences of schizophrenia risk variant rs7914558 at CNNM2 on brain structure. The British Journal of Psychiatry 206(4), 343-344.

Olabi, B., Ellison-Wright, I., McIntosh, A.M., Wood, S.J., Bullmore, E., Lawrie, S.M., 2011. Are there progressive brain changes in schizophrenia? A meta-analysis of structural magnetic resonance imaging studies. Biol Psychiatry 70(1), 88-96.

Owen, M.J., Sawa, A., Mortensen, P.B., 2016. Schizophrenia. Lancet 388(10039), 86-97.

Paul, C.A., Beltz, B., Berger-Sweeney, J., 2008. The nissl stain: a stain for cell bodies in brain sections. CSH Protoc 2008, pdb prot4805.

Peralta, V., Cuesta, M.J., 2011. Eugen Bleuler and the schizophrenias: 100 years after. Schizophr Bull 37(6), 1118-1120.

Perry, T.L., Kish, S.J., Buchanan, J., Hansen, S., 1979. Gamma-aminobutyric-acid deficiency in brain of schizophrenic patients. Lancet 1(8110), 237-239.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29(9), e45.

Pilling, S., Bebbington, P., Kuipers, E., Garety, P., Geddes, J., Orbach, G., et al., 2002. Psychological treatments in schizophrenia: I. Meta-analysis of family intervention and cognitive behaviour therapy. Psychol Med 32(5), 763-782.

Pirinen, M., Lappalainen, T., Zaitlen, N.A., Consortium, G.T., Dermitzakis, E.T., Donnelly, P., et al., 2015. Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics 31(15), 2497-2504.

Pollock, K., Stroemer, P., Patel, S., Stevanato, L., Hope, A., Miljan, E., et al., 2006. A conditionally immortal clonal stem cell line from human cortical neuroepithelium for the treatment of ischemic stroke. Exp Neurol 199(1), 143-155.

Powell, T.R., Murphy, T., Lee, S.H., Duarte, R.R.R., Price, J., Breen, G., et al., 2016. Gene set shows transcriptional changes during human hippocampal neurogenesis and an association with adult hippocampal volume. Manuscript submitted for publication.

Prandovszky, E., Gaskell, E., Martin, H., Dubey, J.P., Webster, J.P., McConkey, G.A., 2011. The Neurotropic Parasite Toxoplasma Gondii Increases Dopamine Metabolism. PLoS One 6(9), e23866.

Pu, J., Schindler, C., Jia, R., Jarnik, M., Backlund, P., Bonifacino, J.S., 2015. BORC, a multisubunit complex that regulates lysosome positioning. Dev Cell 33(2), 176-188.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., et al., 2014. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506(7487), 185-190.

Radewicz, K., Garey, L.J., Gentleman, S.M., Reynolds, R., 2000. Increase in HLA-DR immunoreactive microglia in frontal and temporal cortex of chronic schizophrenics. J Neuropathol Exp Neurol 59(2), 137-150.

Radhakrishnan, R., Wilkinson, S.T., D'Souza, D.C., 2014. Gone to Pot – A Review of the Association between Cannabis and Psychosis. Frontiers in Psychiatry 5, 54.

Ramachandraiah, C.T., Subramaniam, N., Tancer, M., 2009. The story of antipsychotics: Past and present. Indian Journal of Psychiatry 51(4), 324-326.

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., et al., 2014. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17(10), 1418-1428.

Rametti, G., Junque, C., Falcon, C., Bargallo, N., Catalan, R., Penades, R., et al., 2009. A voxel-based diffusion tensor imaging study of temporal white matter in patients with schizophrenia. Psychiatry Res 171(3), 166-176.

Rees, E., Walters, J.T., Chambert, K.D., O'Dushlaine, C., Szatkiewicz, J., Richards, A.L., et al., 2014. CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. Hum Mol Genet 23(6), 1669-1676.

Reynolds, G.P., 2005. The neurochemistry of schizophrenia. Psychiatry 4(10), 21-25.

Richards, A.L., Leonenko, G., Walters, J.T., Kavanagh, D.H., Rees, E.G., Evans, A., et al., 2016. Exome arrays capture polygenic rare variant contributions to schizophrenia. Hum Mol Genet 25(5), 1001-1007.

Riley, B., Thiselton, D., Maher, B.S., Bigdeli, T., Wormley, B., McMichael, G.O., et al., 2010. Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. Mol Psychiatry 15(1), 29-37.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., et al., 2013. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 45(10), 1150-1159.

Rose, E.J., Hargreaves, A., Morris, D., Fahey, C., Tropea, D., Cummings, E., et al., 2014. Effects of a novel schizophrenia risk variant rs7914558 at CNNM2 on brain structure and attributional style. Br J Psychiatry 204(2), 115-121.

Roth, T.L., Lubin, F.D., Funk, A.J., Sweatt, J.D., 2009. LASTING EPIGENETIC INFLUENCE OF EARLY-LIFE ADVERSITY ON THE BDNF GENE. Biological psychiatry 65(9), 760-769.

Roussos, P., Katsel, P., Davis, K.L., Siever, L.J., Haroutunian, V., 2012. A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. Arch Gen Psychiatry 69(12), 1205-1213.

Roussos, P., Mitchell, A.C., Voloudakis, G., Fullard, J.F., Pothula, V.M., Tsang, J., et al., 2014. A role for noncoding variation in schizophrenia. Cell Rep 9(4), 1417-1429.

Roux, P.P., Blenis, J., 2004. ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. Microbiol Mol Biol Rev 68(2), 320-344.

Schizophrenia Commission, 2012. The abandoned illness: a report from the Schizophrenia Commission. Rethink Mental Illness London.

Schizophrenia Psychiatric Genome-Wide Association Study, C., 2011. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 43(10), 969-976.

Schizophrenia Working Group of the PGC, 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511(7510), 421-427.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7), 671-675.

Schreck, R.R., Disteche, C.M., 2001. Chromosome banding techniques. Curr Protoc Hum Genet Chapter 4, Unit4 2.

Schunkert, H., Konig, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., et al., 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 43(4), 333-338.

Schwab, S.G., Kusumawardhani, A.A., Dai, N., Qin, W., Wildenauer, M.D., Agiananda, F., et al., 2013. Association of rs1344706 in the ZNF804A gene with schizophrenia in a case/control sample from Indonesia. Schizophr Res 147(1), 46-52.

Seeman, P., Lee, T., 1975. Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. Science 188(4194), 1217-1219.

Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., et al., 2016. Schizophrenia risk from complex variation of complement component 4. Nature 530(7589), 177-183.

Selemon, L.D., Rajkowska, G., Goldman-Rakic, P.S., 1995. Abnormally high neuronal density in the schizophrenic cortex. A morphometric analysis of prefrontal area 9 and occipital area 17. Arch Gen Psychiatry 52(10), 805-818; discussion 819-820.

Selemon, L.D., Goldman-Rakic, P.S., 1999. The reduced neuropil hypothesis: a circuit based model of schizophrenia. Biol Psychiatry 45(1), 17-25.

Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., et al., 2008. Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. PLoS Genet 4(2), e1000006.

Sharma, M., Kruger, R., Gasser, T., 2014. From genome-wide association studies to next-generation sequencing: lessons from the past and planning for the future. JAMA Neurol 71(1), 5-6.

Sharma, R.P., Grayson, D.R., Gavin, D.P., 2008. Histone deactylase 1 expression is increased in the prefrontal cortex of schizophrenia subjects: analysis of the National Brain Databank microarray collection. Schizophr Res 98(1-3), 111-117.

Shi, Y., Kirwan, P., Smith, J., Robinson, H.P., Livesey, F.J., 2012. Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat Neurosci 15(3), 477-486, S471.

Shorter, K.R., Miller, B.H., 2015. Epigenetic mechanisms in schizophrenia. Prog Biophys Mol Biol 118(1-2), 1-7.

Silva, N.M., Rodrigues, C.V., Santoro, M.M., Reis, L.F., Alvarez-Leite, J.I., Gazzinelli, R.T., 2002. Expression of indoleamine 2,3-dioxygenase, tryptophan degradation, and kynurenine formation during in vivo infection with Toxoplasma gondii: induction by endogenous gamma interferon and requirement of interferon regulatory factor 1. Infect Immun 70(2), 859-868.

Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., McRae, J., et al., 2016. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci 19(4), 571-577.

Smith, R.M., Webb, A., Papp, A.C., Newman, L.C., Handelman, S.K., Suhy, A., et al., 2013. Whole transcriptome RNA-Seq allelic expression in human brain. BMC Genomics 14, 571.

Song, H., Karashima, E., Hamlyn, J.M., Blaustein, M.P., 2014. Ouabain-digoxin antagonism in rat arteries and neurones. J Physiol 592(5), 941-969.

Srivastava, D.P., Woolfrey, K.M., Penzes, P., 2011. Analysis of dendritic spine morphology in cultured CNS neurons. J Vis Exp(53), e2794.

Stefansson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., Ghosh, S., et al., 2002. Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 71(4), 877-892.

Stephan, K.E., Friston, K.J., Frith, C.D., 2009. Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. Schizophr Bull 35(3), 509-527.

Straub, R.E., Jiang, Y., MacLean, C.J., Ma, Y., Webb, B.T., Myakishev, M.V., et al., 2002. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. Am J Hum Genet 71(2), 337-348.

Stuiver, M., Lainez, S., Will, C., Terryn, S., Gunzel, D., Debaix, H., et al., 2011. CNNM2, encoding a basolateral protein required for renal Mg2+ handling, is mutated in dominant hypomagnesemia. Am J Hum Genet 88(3), 333-343.

Sui, J., Pearlson, G.D., Du, Y., Yu, Q., Jones, T.R., Chen, J., et al., 2015. In search of multimodal neuroimaging biomarkers of cognitive deficits in schizophrenia. Biol Psychiatry 78(11), 794-804.

Sullivan, P.F., Kendler, K.S., Neale, M.C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 60(12), 1187-1192.

Sullivan, P.F., 2008. Schizophrenia genetics: the search for a hard lead. Curr Opin Psychiatry 21(2), 157-160.

Sumi, D., Himeno, S., 2012. Role of arsenic (+3 oxidation state) methyltransferase in arsenic metabolism and toxicity. Biol Pharm Bull 35(11), 1870-1875.

Sun, D., van Erp, T.G., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., et al., 2009. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. Biol Psychiatry 66(11), 1055-1060.

Takase, K., Tamagaki, C., Okugawa, G., Nobuhara, K., Minami, T., Sugimoto, T., et al., 2004. Reduced white matter volume of the caudate nucleus in patients with schizophrenia. Neuropsychobiology 50(4), 296-300.

Takata, A., Iwayama, Y., Fukuo, Y., Ikeda, M., Okochi, T., Maekawa, M., et al., 2013. A population-specific uncommon variant in GRIN3A associated with schizophrenia. Biol Psychiatry 73(6), 532-539.

Takeuchi, F., Isono, M., Yamamoto, K., Yokota, M., Akiyama, K., Katsuya, T., et al., 2015. Heterogeneous Effects of Association Between Blood Pressure Loci and Coronary Artery Disease in East Asian Individuals. Circulation Journal 79(4), 830-838.

Talati, P., Rane, S., Kose, S., Blackford, J.U., Gore, J., Donahue, M.J., et al., 2014. Increased hippocampal CA1 cerebral blood volume in schizophrenia. NeuroImage : Clinical 5, 359-364.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., et al., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell 146(6), 1016-1028.

Tansey, K.E., Rees, E., Linden, D.E., Ripke, S., Chambert, K.D., Moran, J.L., et al., 2016. Common alleles contribute to schizophrenia in CNV carriers. Mol Psychiatry 21, 1085-1089.

Tao, R., Cousijn, H., Jaffe, A.E., Burnet, P.W., Edwards, F., Eastwood, S.L., et al., 2014. Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder: a novel transcript fetally regulated by the psychosis risk variant rs1344706. JAMA Psychiatry 71(10), 1112-1120.

Thomas, G.M., Huganir, R.L., 2004. MAPK cascade signalling and synaptic plasticity. Nat Rev Neurosci 5(3), 173-183.

Torrey, E.F., Bartko, J.J., Yolken, R.H., 2012. Toxoplasma gondii and other risk factors for schizophrenia: an update. Schizophr Bull 38(3), 642-647.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., et al., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28(5), 511-515.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., et al., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3), 562-578.

Tyler, C.R., Allan, A.M., 2014. The Effects of Arsenic Exposure on Neurological and Cognitive Dysfunction in Human and Rodent Studies: A Review. Current Environmental Health Reports 1(2), 132-147.

Udler, M., Maia, A.T., Cebrian, A., Brown, C., Greenberg, D., Shah, M., et al., 2007. Common germline genetic variation in antioxidant defense genes and survival after diagnosis of breast cancer. J Clin Oncol 25(21), 3015-3023.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al., 2015. Proteomics. Tissue-based map of the human proteome. Science 347(6220), 1260419.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., et al., 2012. Primer3--new capabilities and interfaces. Nucleic Acids Res 40(15), e115.

van Rossum, J.M., 1966. The significance of dopamine-receptor blockade for the mechanism of action of neuroleptic drugs. Arch Int Pharmacodyn Ther 160(2), 492-494.

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., et al., 2012. Childhood Adversities Increase the Risk of Psychosis: A Meta-analysis of Patient-Control, Prospective- and Cross-sectional Cohort Studies. Schizophrenia Bulletin.

Vassos, E., Collier, D.A., Holden, S., Patch, C., Rujescu, D., St Clair, D., et al., 2010. Penetrance for copy number variants associated with schizophrenia. Hum Mol Genet 19(17), 3477-3481.

Vempati, U.D., Chung, C., Mader, C., Koleti, A., Datar, N., Vidovic, D., et al., 2014. Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS). J Biomol Screen 19(5), 803-816.

Walker, M.A., Highley, J.R., Esiri, M.M., McDonald, B., Roberts, H.C., Evans, S.P., et al., 2002. Estimated neuronal populations and volumes of the hippocampus and its subfields in schizophrenia. Am J Psychiatry 159(5), 821-828.

Wang, J., Duncan, D., Shi, Z., Zhang, B., 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 41(Web Server issue), W77-83.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics 10(1), 57-63.

Weinberger, D.R., 1987. Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 44(7), 660-669.

Weirich, G., Mengele, K., Yfanti, C., Gkazepis, A., Hellmann, D., Welk, A., et al., 2008. Immunohistochemical evidence for ubiquitous distribution of metalloendoprotease insulin-degrading enzyme (IDE; insulysin) in human non-malignant tissues and tumor cell lines. Biological chemistry 389(11), 1441-1445.

Wen, W., Zheng, W., Okada, Y., Takeuchi, F., Tabara, Y., Hwang, J.Y., et al., 2014. Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. Hum Mol Genet 23(20), 5492-5504.

Whitehead, K.A., Langer, R., Anderson, D.G., 2009. Knocking down barriers: advances in siRNA delivery. Nat Rev Drug Discov 8(2), 129-138.

Williams, H.J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., et al., 2011. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. Mol Psychiatry 16(4), 429-441.

Wong, A.H., Van Tol, H.H., 2003. Schizophrenia: from phenomenology to neurobiology. Neurosci Biobehav Rev 27(3), 269-306.

Wong, J., Hyde, T.M., Cassano, H.L., Deep-Soboslay, A., Kleinman, J.E., Weickert, C.S., 2010. Promoter specific alterations of brain-derived neurotrophic factor mRNA in schizophrenia. Neuroscience 169(3), 1071-1084.

Wright, I.C., Rabe-Hesketh, S., Woodruff, P.W., David, A.S., Murray, R.M., Bullmore, E.T., 2000. Meta-analysis of regional brain volumes in schizophrenia. Am J Psychiatry 157(1), 16-25.

Yang, H., Liu, J., Sui, J., Pearlson, G., Calhoun, V.D., 2010. A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining both Improves Classification of Schizophrenia. Frontiers in Human Neuroscience 4, 192.

Yao, J.K., Dougherty, G.G., Reddy, R.D., Keshavan, M.S., Montrose, D.M., Matson, W.R., et al., 2010. Homeostatic Imbalance of Purine Catabolism in First-Episode Neuroleptic-Naïve Patients with Schizophrenia. PLoS One 5(3), e9508.

Zhang, R., Lu, S.M., Qiu, C., Liu, X.G., Gao, C.G., Guo, T.W., et al., 2011. Population-based and family-based association studies of ZNF804A locus and schizophrenia. Mol Psychiatry 16(4), 360-361.

Zhang, R., Yan, J.D., Valenzuela, R.K., Lu, S.M., Du, X.Y., Zhong, B., et al., 2012. Further evidence for the association of genetic variants of ZNF804A with schizophrenia and a meta-analysis for genome-wide significance variant rs1344706. Schizophr Res 141(1), 40-47.

Zhou, Y., Fan, L., Qiu, C., Jiang, T., 2015. Prefrontal cortex and the dysconnectivity hypothesis of schizophrenia. Neurosci Bull 31(2), 207-219.

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., et al., 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48(5), 481-487.

Appendices

Appendix 1. Duarte et al. (2016). Paper on the identification of the putative schizophrenia risk mechanisms on chromosome 10q24.

RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART B
Neuropsychiatric Genetics

Genome-Wide Significant Schizophrenia Risk Variation on Chromosome 10q24 Is Associated With Altered *cis*-Regulation of *BORCS7*, *AS3MT*, and *NT5C2* in the Human Brain

Rodrigo R. R. Duarte,[1] Claire Troakes,[1] Matthew Nolan,[1] Deepak P. Srivastava,[1] Robin M. Murray,[2] and Nicholas J. Bray[1,3]*

[1]Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[3]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, United Kingdom

Manuscript Received: 5 August 2015; Manuscript Accepted: 7 March 2016

Chromosome 10q24.32-q24.33 is one of the most robustly supported risk loci to emerge from genome-wide association studies (GWAS) of schizophrenia. However, extensive linkage disequilibrium makes it difficult to distinguish the actual susceptibility gene(s) at the locus, limiting its value for improving biological understanding of the condition. In the absence of coding changes that can account for the association, risk is likely conferred by altered regulation of one or more genes in the region. We, therefore, used highly sensitive measures of allele-specific expression to assess *cis*-regulatory effects associated with the two best-supported schizophrenia risk variants (SNP rs11191419 and indel ch10_104957618_I/rs202213518) on the primary positional candidates *BORCS7*, *AS3MT*, *CNNM2*, and *NT5C2* in the human brain. Heterozygosity at rs11191419 was associated with increased allelic expression of *BORCS7* and *AS3MT* in the fetal and adult brain, and with reduced allelic expression of *NT5C2* in the adult brain. Heterozygosity at ch10_104957618_I was associated with reduced allelic expression of *NT5C2* in both the fetal and adult brain. Comparisons between cDNA ratios in heterozygotes and homozygotes for the risk alleles indicated that *cis*-effects on *NT5C2* expression in the adult dorsolateral prefrontal cortex could be largely accounted for by genotype at these two risk variants. While not excluding effects on other genes in the region, this study implicates altered neural expression of *BORCS7*, *AS3MT*, and *NT5C2* in susceptibility to schizophrenia arising from genetic variation at the chromosome 10q24 locus.

Key words: GWAS; gene expression; functional genetics; allele-specific expression

How to Cite this Article:
Duarte RRR, Troakes C, Nolan M, Srivastava DP, Murray RM, Bray NJ. 2016. Genome-Wide Significant Schizophrenia Risk Variation on Chromosome 10q24 Is Associated With Altered *cis*-Regulation of *BORCS7*, *AS3MT*, and *NT5C2* in the Human Brain.
Am J Med Genet Part B 171B:806–814.

INTRODUCTION

Chromosome 10q24.32-q24.33 is one of the best-supported genetic risk loci to arise from large-scale genome-wide association studies (GWAS) of schizophrenia [Schizophrenia Psychiatric

Conflicts of interest: The authors declare no conflicts of interest.
Grant sponsor: Science Without Borders (Brazil); Grant number: CAPES BEX 1279-13-0; Grant sponsor: Medical Research Council (UK) Research Grant; Grant number: G0802166; Grant sponsor: Joint MRC/Wellcome Trust; Grant number: 099175/Z/12/Z.
*Correspondence to:
Nicholas Bray, Ph.D., MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, United Kingdom.
E-mail: brayn3@cardiff.ac.uk
Article first published online in Wiley Online Library (wileyonlinelibrary.com): 00 Month 2016
DOI 10.1002/ajmg.b.32445

Genome-Wide Association Study Consortium, 2011; Ripke et al., 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014]. Variation at this locus also exhibits genome-wide significant association with the five disorders included in the Psychiatric Genomics Consortium combined [Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013], suggesting that it increases susceptibility to psychiatric disorders in general. However, like many other loci identified by schizophrenia GWAS, extensive linkage disequilibrium in the region results in association signals spanning multiple genes (Fig. 1), making it difficult to predict the actual susceptibility gene(s) at the locus.

As with the majority of genome-wide significant signals for schizophrenia, the chromosome 10q24 variants exhibiting strongest evidence for association are in non-coding sequence [Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014]. These variants do not appear to index variation influencing protein structure, and are therefore likely to confer risk for schizophrenia through effects on the expression of one or more genes in the region. Measures of allele-specific expression provide a powerful means of assessing such *cis*-regulatory influences, because they allow the level of gene expression from chromosomes carrying the risk and non-risk alleles of a given variant to be compared simultaneously within individual samples [Bray et al., 2003a]. This approach typically makes use of exonic (i.e., expressed) single nucleotide polymorphisms (SNPs) in genes of interest as allele-specific tags, allowing the RNA transcribed from each parental chromosome to be distinguished and relatively quantified in individual heterozygotes [Yan et al., 2002]. A major advantage of this method over traditional expression quantitative trait loci (eQTL) approaches based on total gene expression is that it effectively controls for tissue variables such as RNA quality as well as confounding effects of other genetic and environmental variables, since these influences will usually act on both alleles to the same extent [Bray et al., 2003b].

In order to identify genes that are differentially *cis*-regulated in association with schizophrenia risk variants on chromosome 10q24 (and therefore genes at the locus that potentially confer susceptibility to the disorder), we assessed genotypic effects on the allele-specific expression of the genes encompassed by the strongest schizophrenia association signal: *BORCS7* (formerly *C10ORF32*), *AS3MT*, *CNNM2*, and *NT5C2* (Fig. 1). As *cis*-effects on gene expression can be specific to developmental stage [Hill and Bray, 2012; Tao et al., 2014] and brain region [Buonocore et al., 2010; Gibbs et al., 2010; Ramasamy et al., 2014], we examined effects in the human fetal brain as well as in three adult brain regions implicated in the pathophysiology of schizophrenia: dorsolateral prefrontal cortex (DLPFC), hippocampus, and caudate.

MATERIALS AND METHODS

Brain Samples

Ethical approval for this study was provided by The Joint South London and Maudsley and The Institute of Psychiatry NHS Research Ethics Committee (REF: PNM/12/13-102). Post-mortem human brain tissue from 116 unrelated adults (mean age at death: 72 years; range: 18–102 years) was obtained from the London Neurodegenerative Diseases Brain Bank (UK). All subjects were free from psychiatric or neurological diagnosis at the time of death. Whole brain from 95 second trimester human fetuses

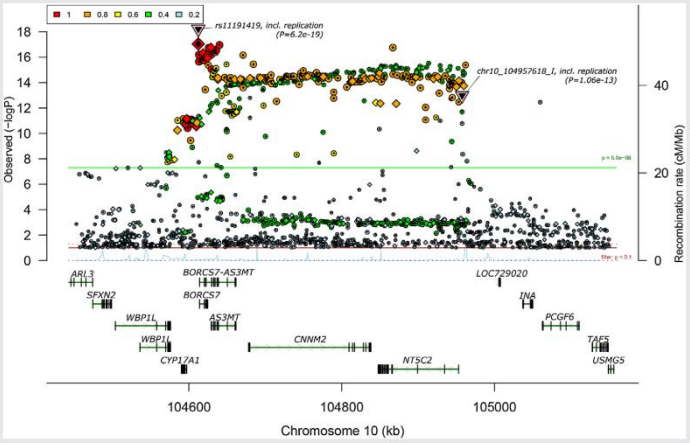


FIG. 1. Genetic association with schizophrenia in a region of strong linkage disequilibrium on chromosome 10q24.32-q24.33. Plot generated by Ricopili [http://www.broadinstitute.org/mpg/ricopili/] using the PGC_SCZ52_may13 dataset described in the Schizophrenia Working Group of the Psychiatric Genomics Consortium [2014] study. Positions of rs11191419 and ch10_104957618_I are indicated by triangles. The threshold of genome-wide significance ($P < 5 \times 10^{-8}$) is indicated by a green horizontal line. Color key indicates r^2 between the variant at the locus showing most significant association with schizophrenia (rs11191419) and other variants in the region. [Color figure can be seen in the online version of this article, available at http://wileyonlinelibrary.com/journal/ajmgb]

(13–23 post-conception weeks) was provided by the MRC—Wellcome Trust Human Developmental Biology Resource (UK). The demographics of samples assayed for each candidate gene are provided in Supplementary Table S1. Genomic DNA was initially extracted from all samples using standard phenol/chloroform procedures, and was used to genotype for the schizophrenia risk variants and the exonic SNPs used to assay the allele-specific expression of each gene. Total RNA was extracted from each brain sample using Tri-Reagent (Life Technologies, Paisley, UK), according to manufacturer's instructions. RNA samples were treated with TURBO DNase (Life Technologies) prior to reverse transcription and did not yield a PCR product in the absence of a reverse transcription step. Approximately 1 μg of total RNA was reverse transcribed with SuperScript III and random decamers (Life Technologies). Resulting cDNA was diluted 1:7 prior to use.

Genotyping

In order to identify heterozygotes informative for the allele-specific expression assays, all samples were initially genotyped for the exonic (expressed) SNPs rs4917985, rs1046778, rs2275271, and rs3740387, tagging *BORCS7*, *AS3MT*, *CNNM2*, and *NT5C2* transcripts, respectively. Genotyping of exonic SNPs was performed by single base primer extension using SNaPshot® chemistry (Life Technologies). All 116 adult and 95 fetal samples were also genotyped for schizophrenia risk SNP rs11191419 and indel ch10_104957618_I (rs202213518). These were selected for study because they are the two independent ($r^2 < 0.2$) variants at the chromosome 10q24 locus showing most significant association with schizophrenia (rs11191419, $P = 6.2 \times 10^{-19}$; ch10_104957618_I, $P = 1.06 \times 10^{-13}$) in the largest GWAS of the disorder published to date [Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014] (Fig. 1). SNP rs11191419 was genotyped by single base primer extension using SNaPshot® chemistry (Life Technologies). The ch10_104957618_I indel was genotyped by Sanger sequencing using BigDye Terminator v3.1 (Life Technologies) in the forward and reverse direction. Primer sequences are provided in Supplementary Table S2. We had previously genotyped all adult samples for rs7085104 and rs11191580, two chromosome 10q24 SNPs reported as genome-wide significant in earlier GWAS of schizophrenia [Schizophrenia Psychiatric Genome-Wide Association Study Consortium, 2011; Ripke et al., 2013], using SNaPshot® primer extension (Life Technologies). SNP rs7085104 is in strong linkage disequilibrium (LD) with rs11191419 in our samples ($r^2 = 0.79$), while SNP rs11191580 is in strong LD with ch10_104957618_I ($r^2 = 0.82$), suggesting that they index the same functional risk variation. There was no significant ($P < 0.05$) deviation from Hardy-Weinberg equilibrium in the genotype distribution of any of the genotyped variants.

Assessment of Allele-Specific Expression

To investigate variable *cis*-effects on the expression of genes at the chromosome 10q24 locus, we used a common exonic SNP in each gene to distinguish the RNA transcribed from each chromosomal copy. Expressed SNPs rs4917985, rs1046778, rs2275271, and rs3740387 were used to tag RNA transcripts for *BORCS7*, *AS3MT*, *CNNM2*, and *NT5C2*, respectively. The allele-specific expression assay was performed using brain cDNA alongside the corresponding genomic DNA from the same subjects. Sequences containing the exonic tag SNPs were PCR-amplified using primers based on single exon sequence, each producing the same amplicon from both cDNA and genomic DNA (Supplementary Table S2). Four technical replicates for each cDNA and genomic DNA sample were assayed for each expressed SNP, with one H_2O-negative control on each plate. PCR products were treated with shrimp alkaline phosphatase and exonuclease I (New England Biolabs, Hitchin, UK) to inactivate nucleotides and primers for downstream steps. Alleles of each expressed SNP were discriminated and relatively quantified by SNaPshot® primer extension (Life Technologies) using extension primers detailed in Supplementary Table S2. Reaction products were electrophoresed on an Applied Biosystems 3130xl Genetic Analyzer and peak heights of allele-specific extended primers were determined using GeneMarker software (SoftGenetics, State College, PA). Peak heights representing the relative abundance of each allele were used to calculate an allele ratio for each reaction. Allele ratios were calculated for each expressed SNP by dividing peak height for the expressed allele that is usually in phase with the risk alleles of rs11191419/ch10_104957618_I by the peak height of the expressed allele that is usually in phase with the non-risk alleles of these variants in samples that were heterozygous at the expressed SNP and the risk variant. For each plate, the average allele ratio from all genomic DNA samples was used as a correction factor for all genomic DNA and cDNA allele ratios, since this can be assumed to reflect a perfect 1:1 ratio of the two alleles and can therefore be used to correct for any inequalities in allelic representation specific to the assay [Bray et al., 2003b]. The average of the four corrected allele ratios for genomic DNA and cDNA from each sample was calculated and used for statistical comparisons. Any samples showing poor reproducibility in cDNA allele ratios (standard deviation/mean > 0.25) were excluded from further analyses.

Assessment of Association Between Schizophrenia Risk Alleles and Allele-Specific Expression

Predicted haplotypes between the schizophrenia risk variants (rs11191419 and ch10_104957618_I) and the expressed SNP for each gene were calculated based on combined genotype data from the 116 adult and 95 fetal samples using Haploview 4.2 software [Barrett et al., 2005]. Predicted haplotype frequencies were used to infer phase between the risk alleles of rs11191419/ch10_104957618_I and the alleles of the exonic SNPs, so that the effect of the risk alleles on gene expression (i.e., up- or down-regulation) could be determined, as described previously [Bray et al., 2005]. Specifically, for samples that were heterozygous at both the exonic SNP and a risk variant, we calculated the frequency of the two possible diplotypes constructed from the two alleles of the exonic SNP and the two alleles of the risk variant on the basis of predicted haplotype frequencies and the assumption of Hardy Weinberg equilibrium using the equations:

$$\text{Frequency diplotype 1} = 2 \times \text{frequency haplotype A} \times \text{frequency haplotype B}$$
$$\text{Frequency diplotype 2} = 2 \times \text{frequency haplotype C} \times \text{frequency haplotype D}$$

The probability that an individual who is heterozygous at both the exonic SNP and the risk variant is carrying diplotype 1 (comprising

haplotypes A and B, rather than haplotypes C and D) is, therefore, calculated by dividing the predicted frequency of diplotype 1 by the combined frequency of both possible diplotypes.

Due to strong linkage disequilibrium in the region, the risk alleles of rs11191419 and ch10_104957618_I would nearly always be carried on the same chromosome as one of the alleles of each exonic SNP. As an initial test of whether the risk alleles were associated with a relative increase or decrease in the allelic expression of each gene, we therefore compared cDNA allele ratios in samples that were heterozygous for each risk variant (where the risk alleles would usually be carried on the same chromosome as one of the expressed alleles and the non-risk alleles would be usually carried on the same chromosome as the alternative expressed allele) with the allele ratios observed in genomic DNA (representing a true 1:1 ratio of the two alleles). As a more specific test of whether risk genotype could account for altered *cis*-regulation of each candidate gene, we also compared cDNA allele ratios between risk allele heterozygotes (where any *cis*-regulatory effects of that variant will differ) and homozygotes (where any *cis*-regulatory effects of the variant will be the same), as performed previously [Bray et al., 2003a, 2005; Hill and Bray, 2012]. Due to small numbers of homozygotes for rs11191419 in assayed exonic heterozygotes, these latter analyses were restricted to comparisons between ch10_104957618_I genotypes. All comparisons were performed by *t*-tests using SPSS 22.0 software. Where differences in variance were detected between comparison groups (Levene's test $P < 0.05$), we used *t*-tests that assumed unequal variance. All tests were two-tailed and *P*-values < 0.05 were considered to be significant. As a more stringent measure of the significance of each finding, we additionally applied a Bonferroni correction for the number of tests performed in each analysis. For the comparisions between allele ratios in cDNA from risk allele heterozygotes and those in genomic DNA, we corrected observed *P*-values for 32 tests (assessing the effects of heterozygosity at two risk variants on the allelic expression of four genes in four brain tissues). For the comparisons between cDNA ratios observed in heterozygotes and homozygotes for ch10_104957618_I, we corrected observed *P*-values for 16 tests (assessing the effect of ch10_104957618_I genotype on four genes in four brain tissues).

RESULTS

One hundred and sixteen adult and 95 fetal human brains were initially genotyped for schizophrenia risk SNP rs11191419 and indel ch10_104957618_I (rs202213518), as well as exonic SNPs in *BORCS7* (rs4917985), *AS3MT* (rs1046778), *CNNM2* (rs2275271), and *NT5C2* (rs3740387), which could serve as allele-specific tags for the four candidate genes in heterozygous samples. Frequencies of the schizophrenia risk alleles in our samples approximated those observed in the control sample of the recent Schizophrenia Working Group of the Psychiatric Genomics Consortium [2014] GWAS. The risk (T) allele of rs11191419 had a reported frequency of 0.64 in the GWAS control samples and a frequency of 0.64 in our combined fetal and adult brain samples, while the risk (deletion) allele of ch10_104957618_I had a reported frequency of 0.92 in the GWAS control samples and a frequency of 0.90 in our samples.

Predicted haplotype frequencies in our entire brain collection allowed us to infer phase between the risk alleles of rs11191419/ch10_104957618_I and the alleles of the exonic SNPs used to assess the allelic expression of each candidate gene (see Materials and Methods). The strong linkage disequilibrium (D′) in the region meant that, in individuals who were heterozygous at a risk variant as well as an exonic variant, the risk allele would nearly always be carried on the same chromosome as a particular allele of the exonic SNP. An assessment of whether the risk alleles were associated with a general increase or decrease in the allelic expression of each candidate gene could, therefore, be made by comparing cDNA allele ratios in individuals who were heterozygous for the risk variant with allele ratios observed in genomic DNA (representing the true 1:1 ratio of the two alleles) [e.g., Bray et al., 2003a, 2005].

Allelic expression data at SNP rs4917985, used to tag *BORCS7*, in heterozygotes for rs11191419 and ch10_104957618_I, are shown in Figure 2A. The C-allele of rs4917985 was predicted to be in phase with the risk (T-) allele of rs11191419 on >99% of occasions when the subject was heterozygous at both loci. Allele ratios in cDNA from rs11191419 heterozygotes indicated a mean increase in expression of the allele carried on the same chromosome as the risk allele, relative to that carried with the non-risk allele, in all assayed tissues (DLPFC: 12%, hippocampus: 8%, caudate: 11%, fetal brain 5%). cDNA allele ratios in all assayed tissues differed significantly from allele ratios observed in genomic DNA (all $P < 0.05$). Although *P*-values survived Bonferroni correction for 32 independent tests only in the DLPFC ($P = 0.001$, corrected $P = 0.032$), *BORCS7* cDNA ratios did not differ significantly from those obtained from rs11191419 heterozygotes in any other adult brain region (all $P > 0.05$). The C-allele of rs4917985 was also predicted to be in phase with the risk (deletion) allele of ch10_104957618_I on >99% of occasions when the subject was heterozygous at both loci. However, allele ratios in cDNA from ch10_104957618_I heterozygotes were close to the 1:1 ratio of equal allelic expression, and did not significantly differ from those observed in genomic DNA in any tissue.

Allelic expression data at SNP rs1046778, used to tag *AS3MT*, in heterozygotes for rs11191419 and ch10_104957618_I, are shown in Figure 2B. The T-allele of rs1046778 was predicted to be in phase with the risk (T-) allele of rs11191419 on >94% of occasions when the subject was heterozygous at both loci. As for *BORCS7*, cDNA allele ratios in rs11191419 heterozygotes indicated a mean increase in expression of the *AS3MT* allele carried on the same chromosome as the risk allele, compared to that carried with the non-risk allele, in all assayed tissues (DLPFC: 14%, hippocampus: 36%, caudate: 23%, fetal brain 40%). cDNA allele ratios differed significantly from allele ratios observed in genomic DNA in fetal brain and in adult hippocampus and caudate ($P < 0.05$ in all tissues). Although *P*-values survived Bonferroni correction for 32 independent tests only in fetal brain ($P = 1.12 \times 10^{-5}$, corrected $P = 0.00036$), *AS3MT* cDNA ratios did not differ significantly from those observed in rs11191419 heterozygotes in adult hippocampus or caudate ($P > 0.05$). The T-allele of rs1046778 was predicted to be in phase with the risk (deletion) allele of ch10_104957618_I on >99% of occasions when the subject was heterozygous at both loci. Relative overexpression of the *AS3MT* allele in phase with the risk allele was

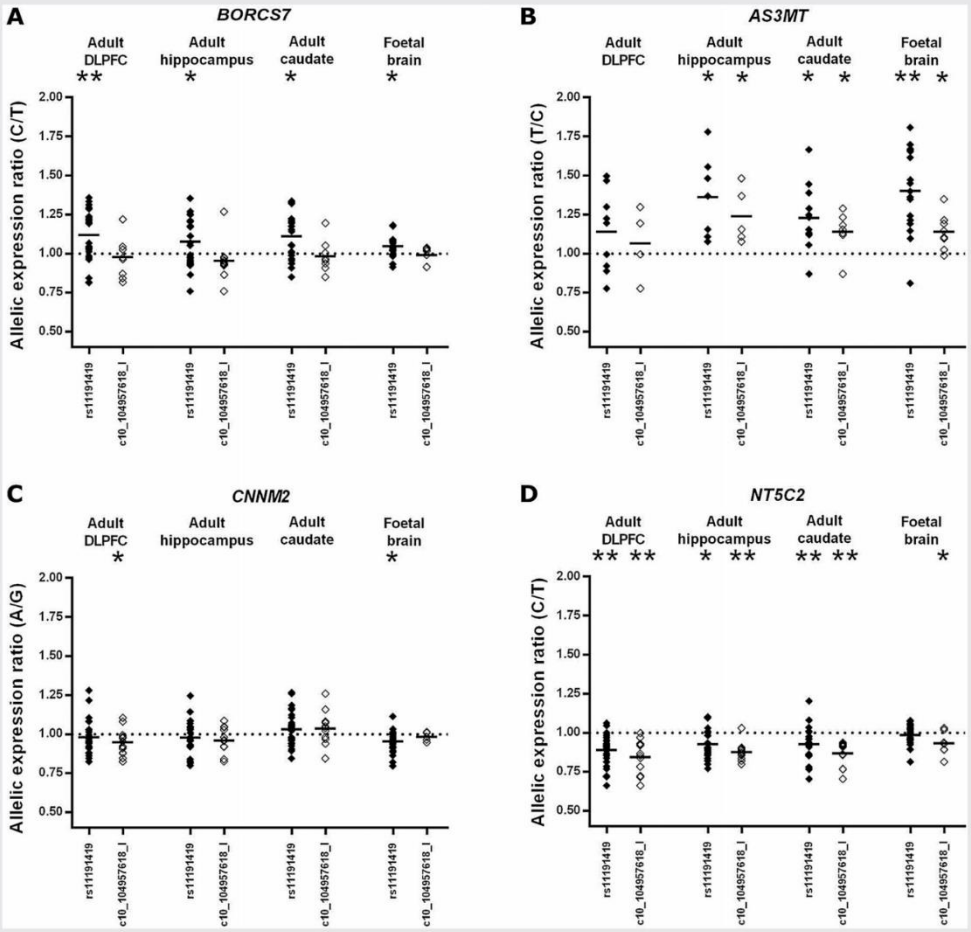


FIG. 2. Allelic expression of *BORCS7* (A), *AS3MT* (B), *CNNM2* (C), and *NT5C2* (D) in heterozygotes for schizophrenia risk variants rs11191419 and ch10_104957618_I. Allelic expression ratios are calculated by dividing measures of the expressed allele that is generally in phase with the schizophrenia risk alleles by measures of the expressed allele that is generally in phase with the non-risk alleles for each susceptibility variant. All raw cDNA ratios are divided by the average allele ratio in genomic DNA (representing the true 1:1 allele ratio) to correct for any inequalities in allelic representation specific to each assay. Data points represent the average of four corrected measures of cDNA allele ratio per sample. Mean corrected cDNA allele ratios are indicated by horizontal lines. The dotted horizontal line indicates the mean genomic DNA (1:1) ratio of the two alleles. Comparisons between cDNA allele ratios in heterozygotes for the risk variants and allele ratios in genomic DNA: $^{*}<0.05$, $^{**}P<0.05$ when Bonferroni corrected for 32 tests.

less pronounced than in rs11191419 heterozygotes (DLPFC: 7%, hippocampus: 24%, caudate: 14%, fetal brain 14%). While allele ratios in ch10_104957618_I heterozygotes differed significantly between cDNA and genomic DNA in hippocampus, caudate and fetal brain (all $P<0.05$), no observation survived Bonferroni correction.

Allelic expression data at SNP rs2275271, used to tag *CNNM2*, in heterozygotes for rs11191419 and ch10_104957618_I, are shown in Figure 2C. The A-allele of rs2275271 was predicted to be in phase with the risk (T-) allele of rs11191419 on >98% of occasions when the subject was heterozygous at both loci. However, unlike cDNA allele ratios in rs11191419 heterozygotes for *BORCS7* and *AS3MT*,

those for *CNNM2* were close to the 1:1 ratio of equal allelic expression, only differing significantly from those in genomic DNA in fetal brain ($P = 0.012$), where a small (5%) relative decrease in the mean expression of the allele generally in phase with the risk allele of rs11191419 was observed. The A-allele of rs2275271 was predicted to be in phase with the risk (deletion) allele of ch10_104957618_I on >99% of occasions when the subject was heterozygous at both loci. Heterozygotes for ch10_104957618_I also displayed little allelic expression imbalance of *CNNM2*, with cDNA allele ratios only differing significantly from those in genomic DNA in the adult DLPFC, where a mean 5% decrease in expression of the allele usually in phase with the risk allele was observed ($P = 0.015$). No allelic expression imbalance of *CNNM2* associated with heterozygosity for the assayed schizophrenia risk variants survived Bonferroni correction.

Allelic expression data at SNP rs3740387, used to tag *NT5C2*, in heterozygotes for rs11191419 and ch10_104957618_I, are shown in Figure 2D. The C-allele of rs3740387 was predicted to be in phase with the risk (T-) allele of rs11191419 on >98% of occasions when the subject was heterozygous at both loci. In rs11191419 heterozygotes, expression of the *NT5C2* allele generally in phase with the risk allele was reduced in all assayed adult brain regions (mean DLPFC: 11%, hippocampus: 7%, caudate: 7%), with cDNA allele ratios differing significantly from those in genomic DNA (all $P < 0.05$). Observed *P*-values survived Bonferroni correction for 32 independent tests in the adult DLPFC ($P = 6.29 \times 10^{-7}$, corrected $P = 2.01 \times 10^{-5}$) and caudate ($P = 0.001$, corrected $P = 0.032$), with a less significant imbalance of allelic expression observed in the hippocampus ($P = 0.003$, corrected $P = 0.083$). The C-allele of rs3740387 was predicted to be in phase with the risk (deletion) allele of ch10_104957618_I on >99% of occasions when the subject was heterozygous at both loci. Expression of the *NT5C2* allele in phase with the ch10_104957618_I risk allele was also reduced in all assayed adult brain regions (mean DLPFC: 15%, hippocampus: 12%, caudate: 13%), with highly significant differences in allele ratios observed between cDNA and genomic DNA that survived Bonferroni correction in all areas (DLPFC: $P = 1.03 \times 10^{-4}$, corrected $P = 0.003$; hippocampus: $P = 2.64 \times 10^{-6}$, corrected $P = 8.46 \times 10^{-5}$; caudate $P = 8.11 \times 10^{-5}$, corrected $P = 0.0025$). Unlike cDNA ratios in rs11191419 heterozygotes, those in ch10_104957618_I heterozygotes also significantly differed from genomic DNA in fetal brain ($P = 0.05$), with the risk allele again associated with reduced *NT5C2* allelic expression, although this latter observation did not survive Bonferroni correction.

Although comparisons between cDNA and gDNA allele ratios in heterozygous risk allele carriers under conditions of high-linkage disequilibrium allow an assessment of whether the risk allele is associated with a general increase or decrease in allelic expression [Bray et al., 2003a, 2005; Hill and Bray, 2012], they do not specifically test whether genotype at the risk variant could directly account for altered *cis*-regulation of the gene. For this, it is necessary to compare cDNA allele ratios in heterozygotes for the risk variant (where any *cis*-regulatory effects of the two alleles will differ) with those in homozygotes for the risk variant (where any *cis*-regulatory effects of the variant will be equal) [Bray et al., 2003a, 2005; Williams et al., 2011; Hill and Bray, 2012]. Small genotype groups precluded such an assessment of *cis*-regulatory effects of rs11191419, but genotype at ch10_104957618_I was associated with significant effects on the allelic expression of *BORCS7* in fetal brain, adult DLPFC, adult hippocampus and adult caudate, *AS3MT* in the fetal brain and adult caudate, and *NT5C2* in the fetal brain, adult DLPFC and adult hippocampus. *P*-values survived Bonferroni correction for 16 tests for *BORCS7* in the adult hippocampus ($P = 0.001$, corrected $P = 0.016$) and fetal brain ($P = 0.002$, corrected $P = 0.032$), *AS3MT* in the fetal brain ($P = 0.002$, corrected $P = 0.032$), and *NT5C2* in the adult DLPFC ($P = 0.003$, corrected $P = 0.048$). Mean cDNA allele ratios at the four candidate genes in ch10_104957618_I heterozygotes and homozygotes are shown in Table I. These analyses showed that the risk allele of ch10_104957618_I is associated with a down-regulation of both *BORCS7* and *AS3MT*, reducing the general overexpression of these genes associated with the risk allele of rs11191419, with less allelic expression imbalance observed in heterozygotes for ch10_104957618_I than in homozygotes. This is consistent with the initial comparisons between allele ratios in cDNA and genomic DNA, which showed less pronounced allelic expression imbalance of *BORCS7* and *AS3MT* in ch10_104957618_I heterozygotes than in rs11191419 heterozygotes. In contrast, the risk alleles of rs11191419 and ch10_104957618_I appear both to be associated with reduced expression of *NT5C2*. Indeed, genotype at rs11191419 and ch10_104957618_I could largely account for observed allelic expression imbalance of *NT5C2* in the adult DLPFC (Fig. 3), where samples that were heterozygous at both risk loci showed a mean 16% reduction in *NT5C2* allelic expression, while homozygotes at both risk loci displayed cDNA allele ratios close to the genomic 1:1 ratio (comparison between cDNA ratios in heterozygotes and homozygotes at both loci: $P = 0.007$).

DISCUSSION

Variants on chromosome 10q24.32-q24.33 exhibit robust association with schizophrenia [Schizophrenia Psychiatric Genome-Wide Association Study Consortium, 2011; Aberg et al., 2013; Ripke et al., 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014], but, like many regions implicated by GWAS, the actual susceptibility genes cannot be easily resolved through genetic data alone. Using a highly sensitive method for assessing variable *cis*-effects on gene expression [Yan et al., 2002; Bray et al., 2003a,b], we have found that several of the principal candidate genes at this locus exhibit altered *cis*-regulation in the developing and adult human brain in association with the most strongly supported schizophrenia risk variants. Largest and most consistent effects were observed on *BORCS7*, *AS3MT*, and *NT5C2*, providing functional support for these as genuine susceptibility genes for schizophrenia.

Our data indicate a complex pattern of *cis*-regulation at the chromosome 10q24 locus. The indel ch10_104957618_I (rs202213518) is located 4,555 bp upstream of the predicted transcription start site of *NT5C2* transcript variant 1 (NM_012229). ENCODE ChIP-seq data indicate that ch10_104957618_I resides in an H3K27ac-marked region that is bound by multiple transcription factors, suggesting direct effects of this variant on *NT5C2* transcription. However, genotype at this variant was also found to influence the allelic expression of *BORCS7* and *AS3MT*, with the

TABLE I. Average Corrected cDNA Allele Ratios at Expressed SNPs in *BORCS7*, *AS3MT*, *CNNM2*, and *NT5C2* According to Genotype at Schizophrenia Risk Variant ch10_104957618_I

Gene (expressed SNP)	Genotype at ch10_104957618_I	Adult DLPFC	Adult hippocampus	Adult caudate	Fetal whole brain
BORCS7 (rs4917985; C/T[a])	Heterozygous	0.99	0.95	0.98	0.99
	Homozygous	1.17	1.12	1.15	1.06
	P het versus hom	**0.006**	**0.001***	**0.004**	**0.002***
AS3MT (rs1046778; T/C)	Heterozygous	1.07	1.24	1.14	1.14
	Homozygous	1.25	1.52	1.31	1.42
	P het versus hom	0.235	0.187	**0.034**	**0.002***
CNNM2 (rs2275271; A/G)	Heterozygous	0.95	0.96	1.04	0.98
	Homozygous	0.99	0.98	1.02	0.95
	P het versus hom	0.095	0.629	0.527	0.342
NT5C2 (rs3740387; C/T)	Heterozygous	0.85	0.88	0.87	0.93
	Homozygous	0.94	0.95	0.93	0.99
	P het versus hom	**0.003***	**0.014**	0.072	**0.016**

[a]Allele ratios at each expressed SNP were calculated by dividing measures of the allele generally in phase with the schizophrenia risk alleles by measures of the allele generally in phase with the non-risk alleles, as indicated.
*P-values surviving Bonferroni correction for 16 tests.
Uncorrected P-values < 0.05 are indicated in bold.

risk allele (in contrast to that of rs11191419) associated with reduced allelic expression of these genes. Similarly, heterozygosity for rs11191419, located within 2 kb of the transcriptional start site of *BORCS7*, was found to be associated with allelic expression imbalance of *BORCS7*, *AS3MT*, and *NT5C2*. This could result from long-range enhancer effects of these variants (or variants in linkage disequilibrium with them), transcriptional interference on adjacent gene expression, or linkage disequilibrium with other functional variants at the chromosome 10q24 locus.

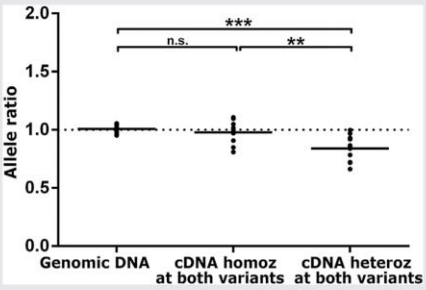


FIG. 3. Allele ratios at expressed *NT5C2* SNP rs3740387 in the DLPFC of adult subjects who are homozygous at both rs11191419 and ch10_104957618_I (6 M, 3 F, average age = 64 years) and in adult subjects who are heterozygous at both variants (9 M, 4 F, average age = 70 years). Data points represent the average of four corrected measures of allele ratio in genomic DNA or cDNA per sample. Mean corrected allele ratios are indicated by horizontal lines. The dotted horizontal line indicates the mean genomic DNA (1:1) ratio of the two alleles. Allele ratios in cDNA from subjects who are homozygous at both risk variants do not significantly differ from those in genomic DNA. Allele ratios in cDNA from subjects who are heterozygous at both risk variants differ significantly from those in genomic DNA and from those in cDNA from homozygotes at both variants. $^{**} < 0.01$, $^{***}P < 0.001$.

It appears that the risk alleles of rs11191419 and ch10_104957618_I have opposing effects on the expression of both *BORCS7* and *AS3MT*, with the risk (T-) allele of rs11191419 associated with increased allelic expression and the risk (deletion) allele of ch10_104957618_I associated with decreased allelic expression of these genes. This is consistent with the risk allele of ch10_104957618_I conferring susceptibility to schizophrenia through effects on a different gene, such as *NT5C2*. However, the risk allele of ch10_104957618_I appears insufficient to fully counteract the increased expression of *AS3MT* associated with the risk allele of rs11191419, with allele ratios in cDNA from ch10_104957618_I heterozygotes remaining significantly higher than the genomic 1:1 ratio in most assayed tissues. In contrast, both risk alleles of rs11191419 and ch10_104957618_I are associated with reduced allelic expression of *NT5C2*, and appear to account for the majority of *cis*-regulatory effects on this gene observed in the adult DLPFC. It is possible that the strong association between rs11191419 and schizophrenia is due to it indexing functional risk variation affecting the regulation of multiple genes at the locus.

Although only some observations survived Bonferroni correction for multiple testing, we urge caution in drawing conclusions as to the relative importance of each finding on the basis of *P*-values alone, since the number of subjects differed between analyses due to differences in expressed allele frequency between candidate genes and the availability of brain tissue from each region. As can be seen in Figure 2, for *BORCS7*, *AS3MT*, and *NT5C2* at least, we observed a general consistency in the effects of risk variant heterozygosity on allelic expression across the brain tissues analyzed.

There are no previous data assessing the impact of ch10_104957618_I genotype on gene expression. However, our findings for rs11191419 appear consistent with existing data generated by eQTL and bioinformatic approaches. The Schizophrenia Working Group of the Psychiatric Genomics Consortium study [2014] included analyses that sought to relate credible GWAS risk variants to genome-wide eQTL data, finding rs11191419 to be in strong linkage disequilibrium ($r^2 = 0.85$) with an eQTL SNP (rs7096169) influencing *AS3MT* expression in blood. Using several brain eQTL datasets, Roussos et al. [2014] identified SNPs influencing *BORCS7 (C10ORF32)*, *AS3MT*, *WBP1L*, and *NT5C2* expression that are in linkage disequilibrium with rs7085104, identified in an earlier GWAS of schizophrenia [Ripke et al., 2013], which we found to be in strong linkage disequilibrium ($r^2 = 0.79$) with rs11191419 in the samples genotyped in the present study. These authors also assessed whether schizophrenia-associated eQTLs were located in predicted *cis*-regulatory elements (CREs); expression of *BORCS7* and *AS3MT* was reported to be influenced by SNPs within individual CRE, while expression of *NT5C2* was associated with SNPs in 14 such elements [Roussos et al., 2014]. Most recently, a genome-wide analysis of DNA methylation QTL in the human fetal brain has indicated that rs7085104 and SNPs in linkage disequilibrium with it are QTL for methylation probes within *AS3MT* [Hannon et al., 2016], consistent with our observation of a large allelic expression imbalance of *AS3MT* in association with rs11191419 heterozygosity in the fetal brain.

Our study is the first to specifically explore effects of chromosome 10q24 schizophrenia risk variants on gene expression in the human fetal brain. Microarray data indicate that *AS3MT* and *NT5C2* are both expressed at a higher level in the prenatal human brain compared to that of the adult [Kang et al., 2011; Birnbaum et al., 2015]. We find that heterozygosity for rs11191419 is associated with particularly pronounced allelic expression imbalance of *AS3MT* in the fetal brain, with an average 40% increase in expression of the *AS3MT* allele that is generally carried on the same chromosome as the risk allele. We find no evidence that *NT5C2* expression is influenced by rs11191419 genotype in fetal brain, but we do observe small effects of ch10_104957618_I genotype on *NT5C2* allelic expression at this early stage of development. These findings would therefore appear consistent with an early neurodevelopmental component to schizophrenia [Murray and Lewis, 1987; Weinberger, 1987], although the observed persistence of effects in the adult brain suggests an ongoing risk mechanism.

A limitation of our study is that it focused on a restricted number of positional candidate genes at the chromosome 10q24.32-q24.33 locus. Although we selected the four candidates flanked by the two best supported risk variants (Fig. 1), extended linkage disequilibrium and the possibility of long-range effects on gene regulation [Sanyal et al., 2012] implicate several other known genes in the region (e.g., *WBP1L*, *CYP17A1*, *INA*, *PCGF6*). In addition, through our use of exonic SNPs that typically tag multiple alternative transcripts of a given gene, we might underestimate *cis*-regulatory effects on individual transcripts, while missing effects on any transcripts that do not include those SNPs. Some of these limitations could be overcome by RNA sequencing, which can be used to measure allele-specific (as well as total) expression of individual transcripts on a genome-wide scale.

The neural functions of the genes implicated in this study remain to be fully elucidated. *BORCS7* encodes BLOC-1-related complex subunit 7 (Diaskedin), part of the recently described BLOC-1-related complex, which has been implicated in lysosomal function and cell migration [Pu et al., 2015]. *AS3MT* encodes arsenic methyltransferase, which has a known role in arsenic metabolism [Sumi and Himeno, 2012], although its functions in the brain are currently unclear. *NT5C2* encodes a cytosolic purine 5′-nucleotidase (cytosolic 5′-nucleotidase II, or cN-II) involved in cellular purine metabolism [Itoh, 2013]. A purinergic hypothesis of schizophrenia has been proposed to explain neurochemical as well as neurodevelopmental aspects of the disorder [Lara and Souza, 2000].

In summary, we have provided an assessment of *cis*-regulatory effects associated with schizophrenia risk variants in a region of extensive linkage disequilibrium on chromosome 10q24. We report altered *cis*-regulation of *BORCS7*, *AS3MT*, and *NT5C2* in association with schizophrenia risk variation, implicating these as genuine schizophrenia susceptibility genes at the locus. Further characterization of these genes in the developing and adult brain is now warranted in order to understand how perturbations in their expression might confer risk for schizophrenia.

ACKNOWLEDGMENTS

This work was funded by a Science without Borders (Brazil) scholarship to R.R.R. Duarte (CAPES BEX 1279-13-0) and a Medical Research Council (UK) Research Grant to N.J. Bray (grant #G0802166). The human fetal material was provided by the Joint MRC/Wellcome Trust (grant #099175/Z/12/Z) Human Developmental Biology Resource (www.hdbr.org). Adult tissue samples were supplied by The London Neurodegenerative Diseases Brain Bank, which receives funding from the MRC and, as part of the Brains for Dementia Research programme, jointly funded by Alzheimer's Research UK and Alzheimer's Society.

REFERENCES

Aberg KA, Liu Y, Bukszar J, McClay JL, Khachane AN, Andreassen OA, Blackwood D, Corvin A, Djurovic S, Gurling H, Ophoff R, Pato CN, Pato MT, Riley B, Webb T, Kendler K, O'Donovan M, Craddock N, Kirov G, Owen M, Rujescu D, St. Clair D, Werge T, Hultman CM, Delisi LE, Sullivan P, van den Oord EJ. 2013. A comprehensive family-based replication study of schizophrenia genes. JAMA Psychiatry 70:573–581.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Birnbaum R, Jaffe AE, Chen Q, Hyde TM, Kleinman JE, Weinberger DR. 2015. Investigation of the prenatal expression patterns of 108 schizophrenia-associated genetic loci. Biol Psychiatry 77:e43–e51.

Bray NJ, Buckland PR, Williams NM, Williams HJ, Norton N, Owen MJ, O'Donovan MC. 2003a. A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. Am J Hum Genet 73:152–161.

Bray NJ, Buckland PR, Owen MJ, O'Donovan MC. 2003b. *Cis*-acting variation in the expression of a high proportion of genes in human brain. Hum Genet 113:149–153.

Bray NJ, Preece A, Williams NM, Moskvina V, Buckland PR, Owen MJ, O'Donovan MC. 2005. Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression. Hum Mol Genet 14:1947–1954.

Buonocore F, Hill MJ, Campbell CD, Oladimeji PB, Jeffries AR, Troakes C, Hortobagyi T, Williams BP, Cooper JD, Bray NJ. 2010. Effects of *cis*-regulatory variation differ across regions of the adult human brain. Hum Mol Genet 19:4490–4496.

Cross-Disorder Group of the Psychiatric Genomics Consortium. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. Lancet 381:1371–1379.

Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, Singleton AB. 2010. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 6:e1000952.

Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, Troakes C, Turecki G, O'Donovan MC, Schalkwyk LC, Bray NJ, Mill J. 2016. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci 19:48–54.

Hill MJ, Bray NJ. 2012. Evidence that schizophrenia risk variation in the ZNF804A gene exerts its effects during fetal brain development. Am J Psychiatry 169:1301–1308.

Itoh R. 2013. Enzymatic properties and physiological roles of cytosolic 5′-nucleotidase II. Curr Med Chem 20:4260–4284.

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G, Guennel T, Shin Y, Johnson MB, Krsnik Z, Mayer S, Fertuzinhos S, Umlauf S, Lisgo SN, Vortmeyer A, Weinberger DR, Mane S, Hyde TM, Huttner A, Reimers M, Kleinman JE, Sestan N. 2011. Spatio-temporal transcriptome of the human brain. Nature 478:483–489.

Lara DR, Souza DO. 2000. Schizophrenia: A purinergic hypothesis. Med Hypotheses 54:157–166.

Murray RM, Lewis SW. 1987. Is schizophrenia a neurodevelopmental disorder? Br Med J 295:681–682.

Pu J, Schindler C, Jia R, Jarnik M, Backlund P, Bonifacino JS. 2015. BORC, a multisubunit complex that regulates lysosome positioning. Dev Cell 33:176–188.

Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, UK Brain Expression Consortium, North American Brain Expression Consortium, Coin L, de Silva R, Cookson MR, Singleton AB, Hardy J, Ryten M, Weale ME. 2014. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17:1418–1428.

Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, Lee SH, Magnusson PK, Sanchez N, Stahl EA, Williams S, Wray NR, Xia K, Bettella F, Borglum AD, Bulik-Sullivan BK, Cormican P, Craddock N, de Leeuw C, Durmishi N, Gill M, Golimbet V, Hamshere ML, Holmans P, Hougaard DM, Kendler KS, Lin K, Morris DW, Mors O, Mortensen PB, Neale BM, O'Neill FA, Owen MJ, Milovancevic MP, Posthuma D, Powell J, Richards AL, Riley BP, Ruderfer D, Rujescu D, Sigurdsson E, Silagadze T, Smit AB, Stefansson H, Steinberg S, Suvisaari J, Tosato S, Verhage M, Walters JT, Multicenter Genetic Studies of Schizophrenia Consortium, Levinson DF, Gejman PV, Kendler KS, Laurent C, Mowry BJ, O'Donovan MC, Owen MJ, Pulver AE, Riley BP, Schwab SG, Wildenauer DB, Dudbridge F, Holmans P, Shi J, Albus M, Alexander M, Campion D, Cohen D, Dikeos D, Duan J, Eichhammer P, Godard S, Hansen M, Lerer FB, Liang KY, Maier W, Mallet J, Nertney DA, Nestadt G, Norton N, O'Neill FA, Papadimitriou GN, Ribble R, Sanders AR, Silverman JM, Walsh D, Williams NM, Wormley B, Psychosis Endophenotypes International Consortium, Arranz MJ, Bakker S, Bender S, Bramon E, Collier D, Crespo-Facorro B, Hall J, Iyegbe C, Jablensky A, Kahn RS, Kalaydjieva L, Lawrie S, Lewis CM, Lin K, Linszen DH, Mata I, McIntosh A, Murray RM, Ophoff RA, Powell J, Rujescu D, Van Os J, Walshe M, Weisbrod M, Wiersma D, Wellcome Trust Case Control Consortium 2, Donnelly P, Barroso I, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin AP, Deloukas P, Duncanson A, Jankowski J, Markus HS, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Trembath RC, Viswanathan AC, Wood NW, Spencer CC, Band G, Bellenguez C, Freeman C, Hellenthal G, Giannoulatou E, Pirinen M, Pearson RD, Strange A, Su Z, Vukcevic D, Donnelly P, Langford C, Hunt SE, Edkins S, Gwilliam R, Blackburn H, Bumpstead SJ, Dronov S, Gillman M, Gray E, Hammond N, Jayakumar A, McCann OT, Liddle J, Potter SC, Ravindrarajah R, Ricketts M, Tashakkori-Ghanbaria A, Waller MJ, Weston P, Widaa S, Whittaker P, Barroso I, Deloukas P, Mathew CG, Blackwell JM, Brown MA, Corvin AP, McCarthy MI, Spencer CC, Bramon E, Corvin AP, O'Donovan MC, Stefansson K, Scolnick E, Purcell S, McCarroll SA, Sklar P, Hultman CM, Sullivan PF. 2013. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 45:1150–1159.

Roussos P, Mitchell AC, Voloudakis G, Fullard JF, Pothula VM, Tsang J, Stahl EA, Georgakopoulos A, Ruderfer DM, Charney A, Okada Y, Siminovitch KA, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Plenge RM, Raychaudhuri S, Fromer M, Purcell SM, Brennand KJ, Robakis NK, Schadt EE, Akbarian S, Sklar P. 2014. A role for noncoding variation in schizophrenia. Cell Rep 9:1417–1429.

Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. Nature 489:109–113.

Schizophrenia Psychiatric Genome-Wide Association Study Consortium. 2011. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 43:969–976.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.

Sumi D, Himeno S. 2012. Role of arsenic (+3 oxidation state) methyltransferase in arsenic metabolism and toxicity. Biol Pharm Bull 35:1870–1875.

Tao R, Cousijn H, Jaffe AE, Burnet PW, Edwards F, Eastwood SL, Shin JH, Lane TA, Walker MA, Maher BJ, Weinberger DR, Harrison PJ, Hyde TM, Kleinman JE. 2014. Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder: A novel transcript fetally regulated by the psychosis risk variant rs1344706. JAMA Psychiatry 71:1112–1120.

Weinberger DR. 1987. Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 44:660–669.

Williams HJ, Norton N, Dwyer S, Moskvina V, Nikolov I, Carroll L, Georgieva L, Williams NM, Morris DW, Quinn EM, Giegling I, Ikeda M, Wood J, Lencz T, Hultman C, Lichtenstein P, Thiselton D, Maher BS, Molecular Genetics of Schizophrenia Collaboration (MGS) International Schizophrenia Consortium (ISC), SGENE-plus GROUP, Malhotra AK, Riley B, Kendler KS, Gill M, Sullivan P, Sklar P, Purcell S, Nimgaonkar VL, Kirov G, Holmans P, Corvin A, Rujescu D, Craddock N, Owen MJ, O'Donovan MC. 2011. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. Mol Psychiatry 16:429–441.

Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2002. Allelic variation in human gene expression. Science 297:1143.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

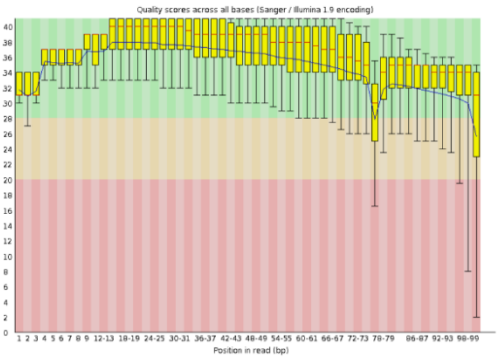
Appendix 2: RNA-seq metrics from raw FASTQ files (FASTQC)

A. DLPFC sample – forward sequencing

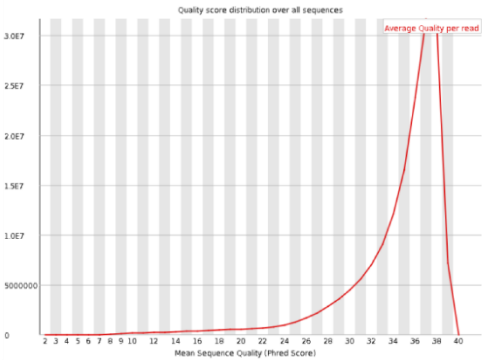
Basic Statistics

Measure	Value
Filename	130_09_DLPFC_l1_ACAGTG_L008_R1_001.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	167943560
Sequences flagged as poor quality	0
Sequence length	100
%GC	44

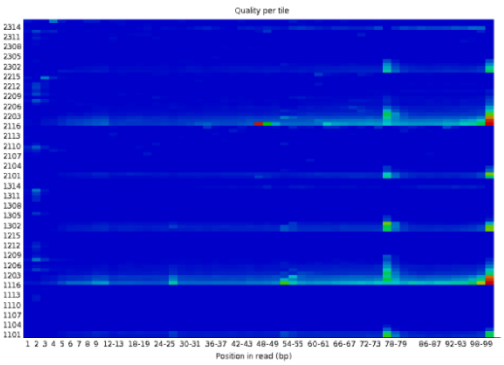
Per base sequence quality



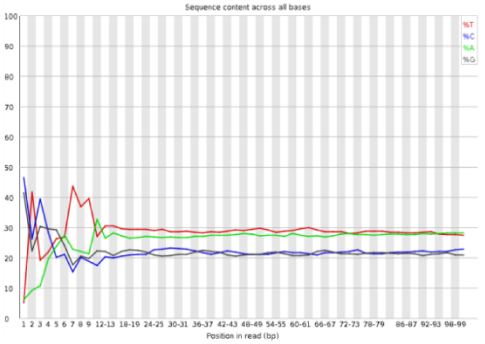
Per sequence quality scores



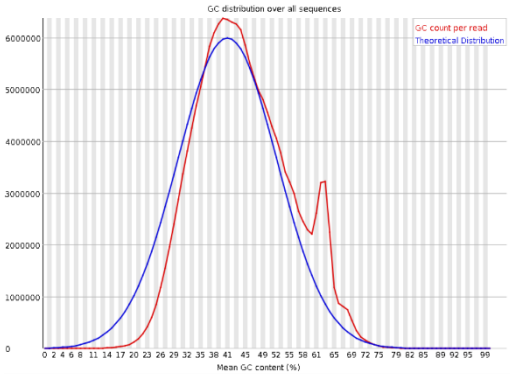
Per tile sequence quality



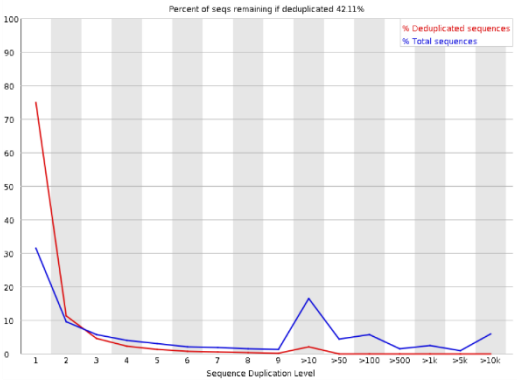
Per base sequence content



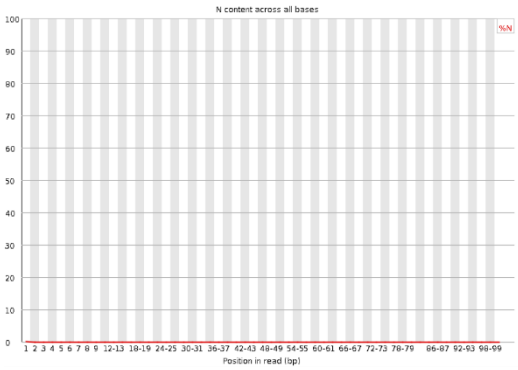
Per sequence GC content



Sequence Duplication Levels



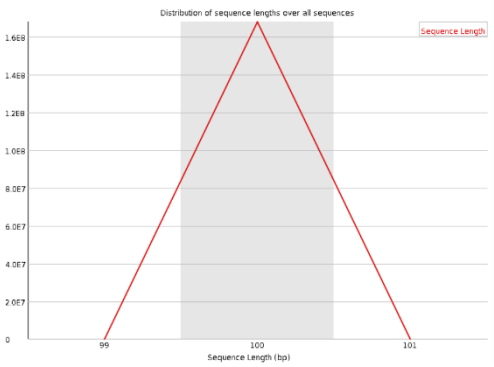
Per base N content



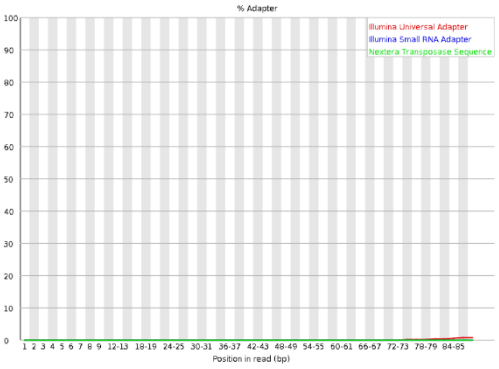
Overrepresented sequences

Sequence	Counts	%	Possible Source
CCCCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATAT	1004245	0.59	No Hit
CCCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATT	742172	0.44	No Hit
CTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGA	672827	0.40	No Hit
CACCCCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCAT	486261	0.28	No Hit
CCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTG	384849	0.22	No Hit
CCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGATG	359267	0.21	No Hit
GCCCTCTTGAACTCTCTCTTCAAAGTTCTTTTCAACTTTCCCTTACGGTA	283307	0.16	No Hit
CTGGAGTCTTGGAAGCTTGACTACCCTACGTTCTCCTACAAATGGACCTT	282061	0.16	No Hit
CTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTTAGGCAACCTGGTGGT	173589	0.10	No Hit

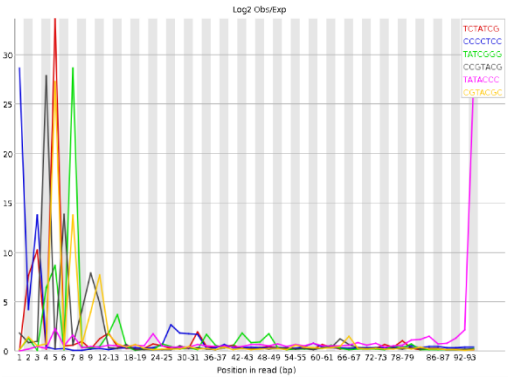
Sequence Length Distribution



Adapter Content



Kmer Content



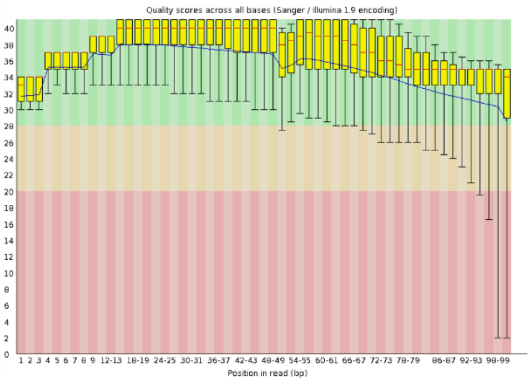
Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
TCTATCG	47525	0.0	33.58045	5
CCCCTCC	420425	0.0	28.671118	1
TATCGGG	55825	0.0	28.663506	7
CCGTACG	14315	0.0	27.871304	4
TATACCC	98855	0.0	27.571558	94
CGTACGC	14505	0.0	27.311827	5
GCCCAGA	258235	0.0	26.897223	94
ACCCCTC	203375	0.0	25.25132	2
GTCGGAT	11035	0.0	24.876507	1
CACCCCT	224355	0.0	24.548891	1
GTCCGGT	15955	0.0	24.494833	1
GTCCCGT	18430	0.0	24.450052	1
CCCGTCT	57700	0.0	24.44889	1
CGTATAC	61890	0.0	23.678951	94
ACGGCGA	13595	0.0	23.474558	94
TTCGCGC	5270	0.0	23.452333	7
CGTCGGA	15045	0.0	23.367718	94
GTCGGGT	24285	0.0	23.344337	1
GTCCGTT	17005	0.0	23.231567	1
GTACGCC	17195	0.0	23.066471	6

B. DLPFC sample – reverse sequencing

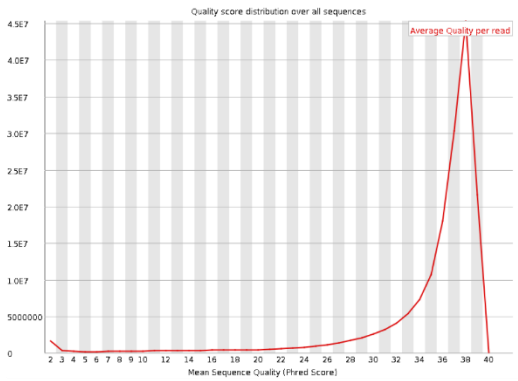
Basic Statistics

Measure	Value
Filename	130_09_DLPFC_I1_ACAGTG_L008_R2_001.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	167943560
Sequences flagged as poor quality	0
Sequence length	100
%GC	43

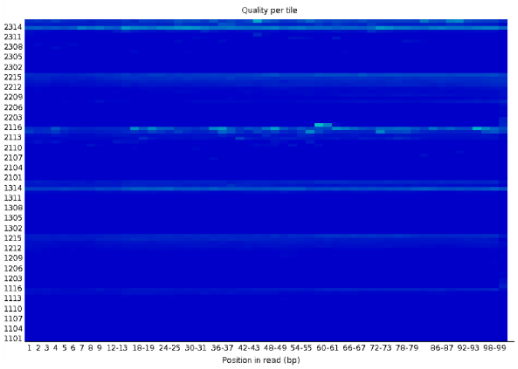
Per base sequence quality



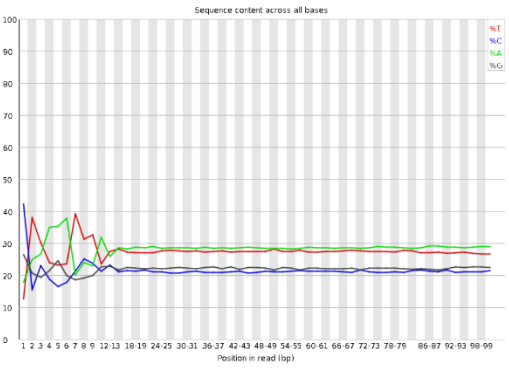
Per sequence quality scores



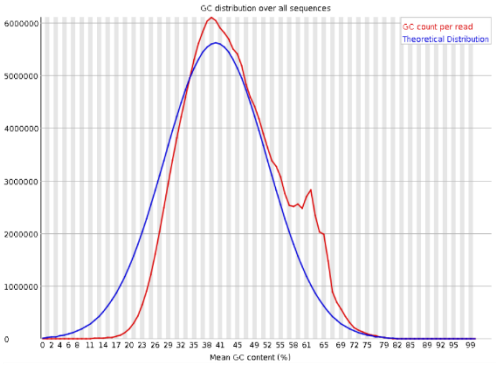
Per tile sequence quality



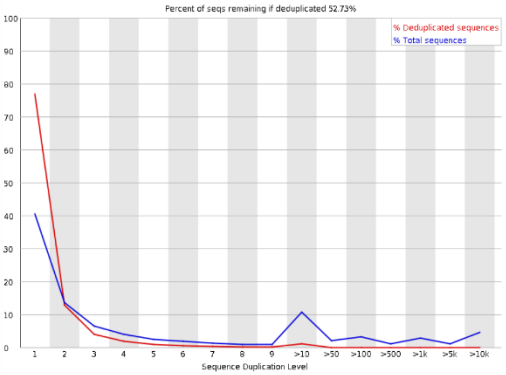
Per base sequence content



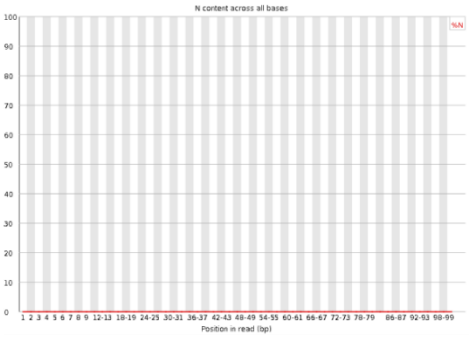
Per sequence GC content



Sequence Duplication Levels



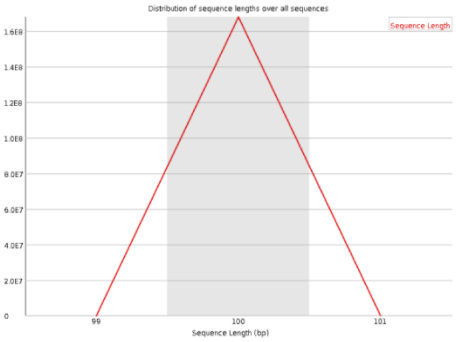
Per base N content



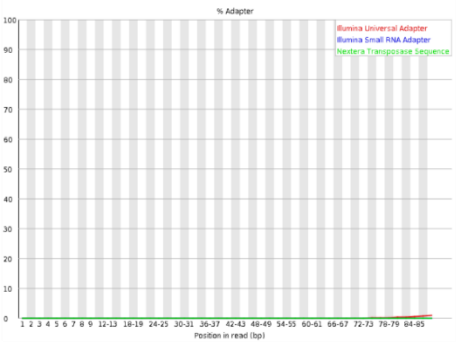
Overrepresented sequences

Sequence	Counts	%	Possible Source
CGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGGATCGCT	356495	0.21	No Hit
CCCAGCTACTCGGGAGGCTGAGGTGGGAGGATCGCTTGAGCCCAGGAGTT	331406	0.19	No Hit
CGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCTGGAGGATCGCT	205236	0.12	No Hit
GGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGGA	199214	0.11	No Hit
CGGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGG	195887	0.11	No Hit
CCCAGCTACTCGGGAGGCTGAGGCTGGAGGATCGCTTGAGTCCAGGAGTT	194916	0.11	No Hit

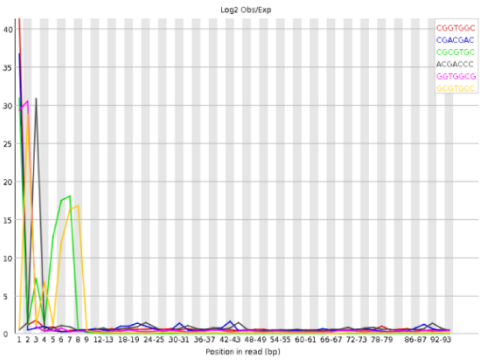
Sequence Length Distribution



Adapter Content



Kmer Content



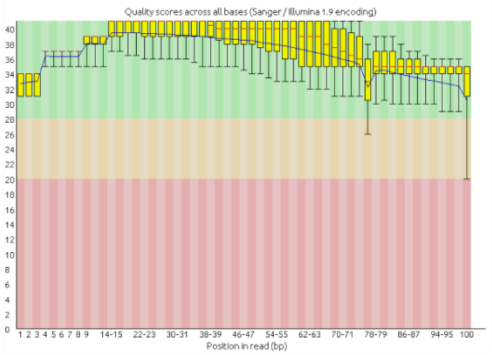
Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CGGTGGC	90690	0.0	41.28851	1
CGACGAC	27970	0.0	36.697357	1
CGCGTGC	186015	0.0	30.935982	1
ACGACCC	33225	0.0	30.810963	3
GGTGGCG	110735	0.0	30.550056	2
GCGTGCC	200210	0.0	28.807835	2
GGCGCGT	119975	0.0	28.096926	5
CGACCCA	36520	0.0	28.00543	4
TGGCGCG	121535	0.0	27.743176	4
ATCGGGG	43545	0.0	26.767424	94
GCGCGTG	126585	0.0	26.63603	6
GTGGCGC	130750	0.0	25.863258	3
CGTGCCT	272425	0.0	21.463268	3
GACGACC	57640	0.0	21.191101	2
GCGACAT	39070	0.0	20.379843	7
CATTCGA	49765	0.0	20.29965	9
CCTGTAG	293900	0.0	19.94333	7
TGCGACA	39660	0.0	19.867794	6
TGTAGTC	294285	0.0	19.841047	9
CCATTCG	51205	0.0	19.837229	8

C. Foetal brain sample – forward sequencing

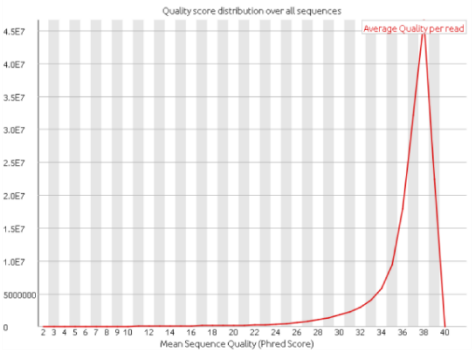
Basic Statistics

Measure	Value
Filename	15533_1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	153548265
Sequences flagged as poor quality	0
Sequence length	100
%GC	45

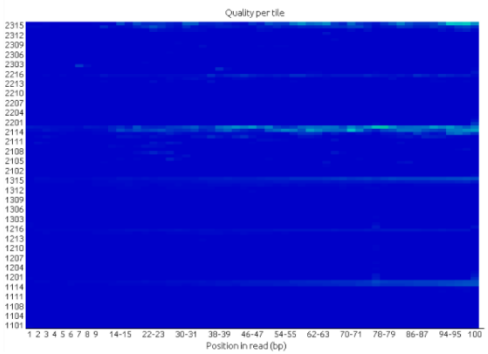
Per base sequence quality



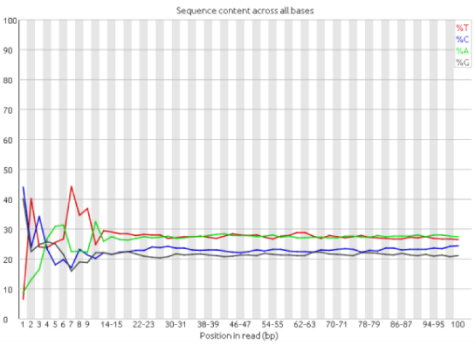
Per sequence quality scores



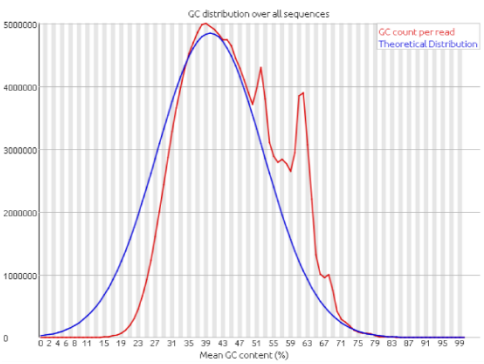
Per tile sequence quality



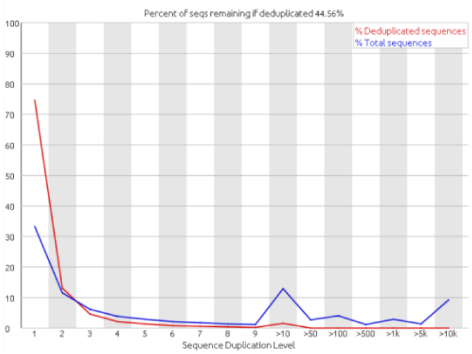
Per base sequence content



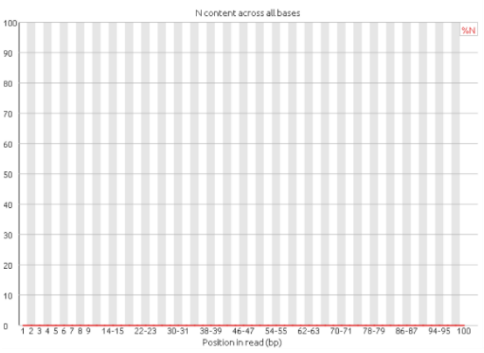
Per sequence GC content



Sequence Duplication Levels



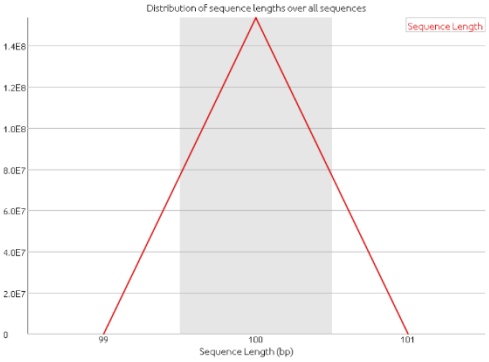
Per base N content



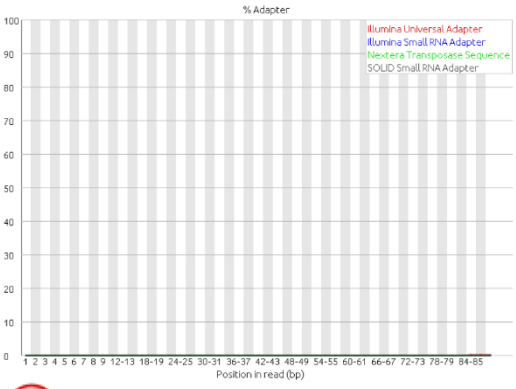
Overrepresented sequences

Sequence	Count	%	Possible Source
CTGGAGTCTTGGAAGCTTGACTACCCTACGTTCTCCTACAAATGGACCTT	773839	0.50	No Hit
CCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGATG	658635	0.42	No Hit
GTCTGGAGTCTTGGAAGCTTGACTACCCTACGTTCTCCTACAAATGGACC	499261	0.32	No Hit
CCCCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATAT	390443	0.25	No Hit
CCCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATT	353245	0.23	No Hit
GGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGATGCCGAA	325958	0.21	No Hit
CTCAGGCTGGAGTGCAGTGGCTATTCACAGGCGCGATCCCACTACTGATC	322134	0.20	No Hit
CTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGA	310035	0.20	No Hit
CCAGGCTGGAGTGCAGTGGCTATTCACAGGCGCGATCCCACTACTGATCA	290180	0.18	No Hit
GTGGCTATTCACAGGCGCGATCCCACTACTGATCAGCACGGGAGTTTTGA	287631	0.18	No Hit
CTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTTAGGCAACCTGGTGGT	270045	0.17	No Hit
CTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGATGC	263486	0.17	No Hit
CAGGCTGGAGTGCAGTGGCTATTCACAGGCGCGATCCCACTACTGATCAG	243640	0.15	No Hit
CCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTG	242193	0.15	No Hit
CTCAGACCGCGTTCTCTCCCTCTCACTCCCCAATACGGAGAGAAGAACGA	228656	0.14	No Hit
GCTCAGGCTGGAGTGCAGTGGCTATTCACAGGCGCGATCCCACTACTGAT	207860	0.13	No Hit
GCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTTAGGCAACCTGGTGG	171001	0.11	No Hit
GGCTGGAGTGCAGTGGCTATTCACAGGCGCGATCCCACTACTGATCAGCA	169646	0.11	No Hit
GGCTATTCACAGGCGCGATCCCACTACTGATCAGCACGGGAGTTTTGACC	162718	0.10	No Hit

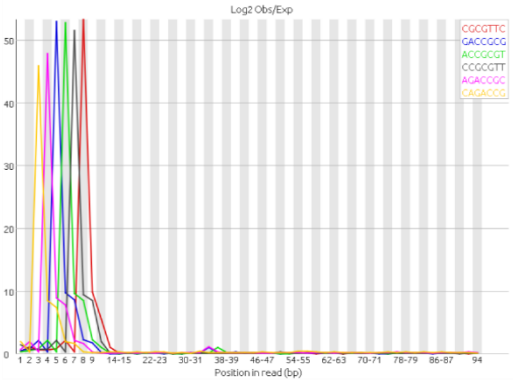
Sequence Length Distribution



Adapter Content



Kmer Content



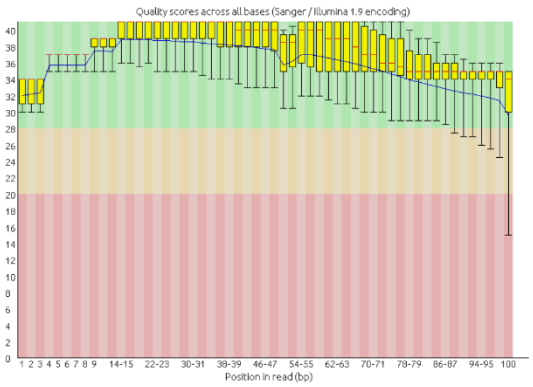
Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CGCGTTC	43365	0.0	53.21231	8
GACCGCG	43695	0.0	53.004036	5
ACCGCGT	43775	0.0	52.789074	6
CCGCGTT	44835	0.0	51.520058	7
AGACCGC	48375	0.0	47.837345	4
CAGACCG	50705	0.0	45.935722	3
GCGTTCT	57100	0.0	40.601772	9
GCCGGCT	62765	0.0	36.26418	94
TATACCC	218270	0.0	34.44924	94
TGGAGTC	269265	0.0	28.577251	2
GGAGTCT	273330	0.0	28.418621	3
GAGTCTT	273245	0.0	28.267504	4
GTCTGGA	189360	0.0	28.024706	1
GTCTTGG	282470	0.0	27.625515	6
AGTCTTG	284750	0.0	27.318495	5
TACGGAG	51795	0.0	26.164751	34-35
TTCGCGC	10240	0.0	25.104836	7
TCTTGGA	317850	0.0	24.841791	7
CGGCATG	43795	0.0	24.810373	3
CAACGCC	54330	0.0	24.692734	82-83

D. Foetal brain sample – reverse sequencing

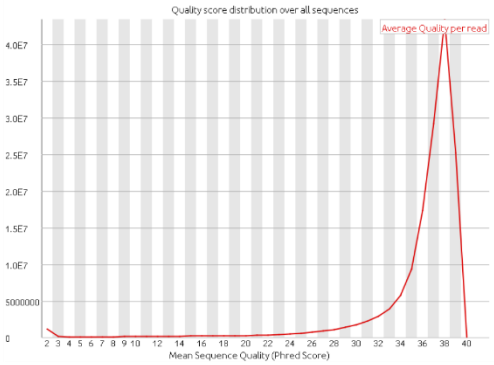
Basic Statistics

Measure	Value
Filename	15533_2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	153548265
Sequences flagged as poor quality	0
Sequence length	100
%GC	45

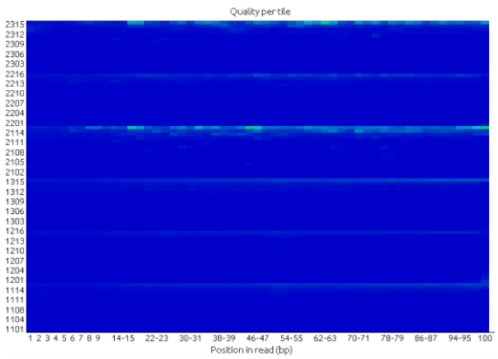
Per base sequence quality



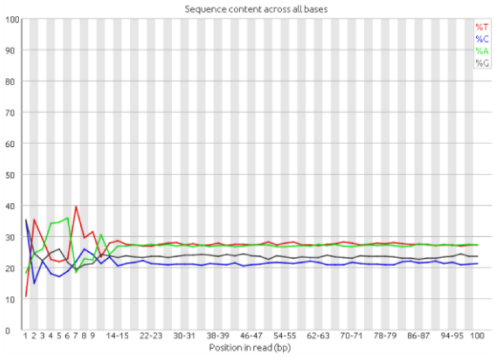
Per sequence quality scores



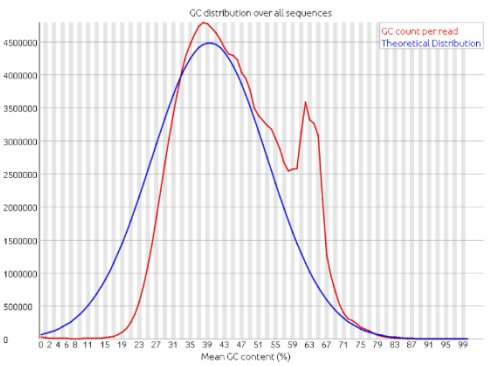
Per tile sequence quality



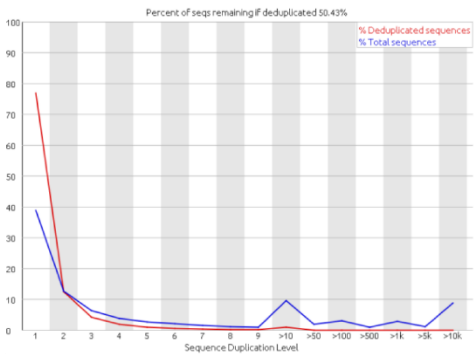
Per base sequence content



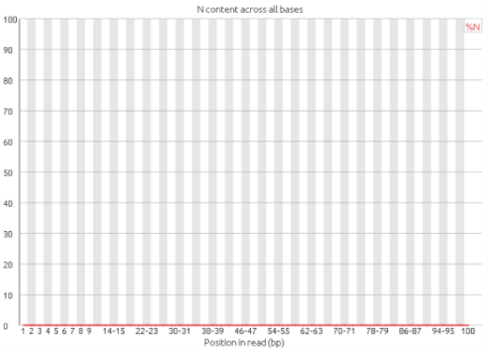
Per sequence GC content



Sequence Duplication Levels



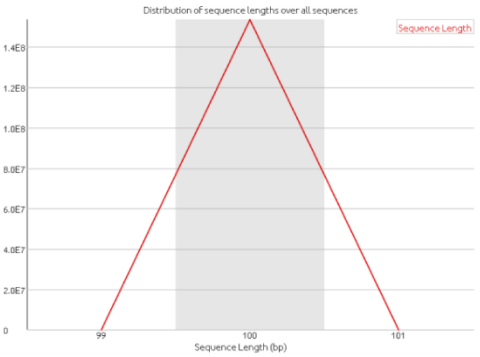
Per base N content



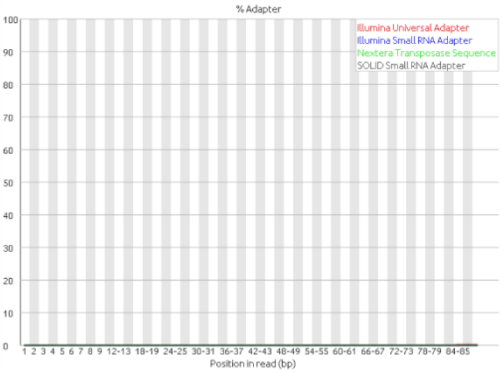
Overrepresented sequences

Sequence	Count	%	Possible Source
GGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGGA	465566	0.30	No Hit
GTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGGAT	424757	0.27	No Hit
GGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCTGGAGGA	332000	0.21	No Hit
CGGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGG	329618	0.21	No Hit
GTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCTGGAGGAT	305020	0.19	No Hit
CCCAGCTACTCGGGAGGCTGAGGTGGGAGGATCGCTTGAGCCCAGGAGTT	287685	0.18	No Hit
GGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGGATCG	258994	0.16	No Hit
CTGGCTGCGACATCTGTCACCCCATTGATCGCCAGGGTTGATTCGGCTGA	253829	0.16	No Hit
CGGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCTGGAGG	235757	0.15	No Hit
CCCAGCTACTCGGGAGGCTGAGGCTGGAGGATCGCTTGAGTCCAGGAGTT	201356	0.13	No Hit
GGGCGATCTGGCTGCGACATCTGTCACCCCATTGATCGCCAGGGTTGATT	201154	0.13	No Hit
GGCGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCTGGAGGATCG	184736	0.12	No Hit
GGGAAGCTCATCAGTGGGGCCACGAGCTGAGTGCGTCCTGTCACTCCACT	167364	0.10	No Hit
CATCTGTCACCCCATTGATCGCCAGGGTTGATTCGGCTGATCTGGCTGGC	165359	0.10	No Hit
CGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGTGGGAGGATCGCT	161475	0.10	No Hit

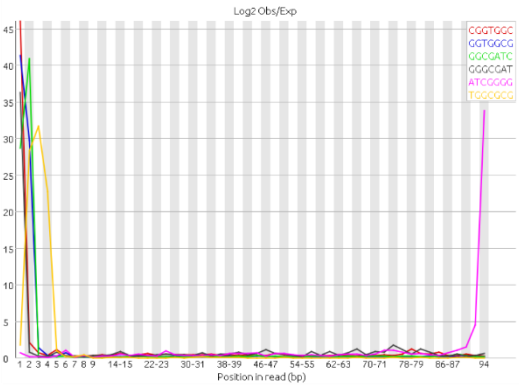
Sequence Length Distribution



Adapter Content



Kmer Content



Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CGGTGGC	126545	0.0	46.000404	1
GGTGGCG	194560	0.0	41.327866	1
GGCGATC	48175	0.0	40.880653	2
GGGCGAT	54775	0.0	36.32799	1
ATCGGGG	58080	0.0	33.79326	94
TGGCGCG	252985	0.0	31.677664	3
GCGATCT	65240	0.0	30.648348	3
CGATCTG	65050	0.0	30.580065	4
GTGGCGC	269490	0.0	29.807318	2
TCACCCG	38425	0.0	28.09769	3
GGCGCGT	288615	0.0	27.916267	4
GCGCGTG	302715	0.0	26.619274	5
CGCGTGC	330170	0.0	24.33892	6
TGCGACA	107905	0.0	23.681417	6
GCGACAT	110085	0.0	23.443907	7
GCGTGCC	345015	0.0	23.382471	7
CTGCGAC	115235	0.0	22.234896	5
GCTGCGA	115690	0.0	22.122904	4
ACCCGGC	51420	0.0	21.774742	5
CGACATC	118205	0.0	21.754333	8

Appendix 3A. Connectivity mapping results for top 200 compounds reversing the NT5C2 knockdown signature in all cell lines available on LincsCloud.

Rank	REVERSING SIGNATURE sig_id	pert_id	pert_iname	pert_type	cell_id	pert_itime	pert_idose	score
1	CSS001_A549_96H:CAGATCC:1	CSS001-CAGATCC	CAGATCC	trt_sh.css	A549	96 h	1 µL	-0.5472
2	KDA006_MCF7_144H:TRCN0000002390:-666	TRCN0000002390	SCYL3	trt_sh	MCF7	144 h	2 µL	-0.546
3	OEB001_PC3_96H:BRDN0000398766:-666	ccsbBroad304_00534	FABP1	trt_oe	PC3	96 h	2 µL	-0.5445
4	CPC014_A375_6H:BRD-K02965346-001-01-8:10	BRD-K02965346	SU-11274	trt_cp	A375	6 h	10 µM	-0.5311
5	DOS014_VCAP_6H:BRD-K53507509-001-01-9:5.05	BRD-K53507509	BRD-K53507509	trt_cp	VCAP	6 h	5 µM	-0.5275
6	OEB001_A375_96H:BRDN0000399999:-666	ccsbBroad304_02889	WWTR1	trt_oe	A375	96 h	2 µL	-0.5265
7	KDD003_PC3_96H:TRCN0000007362:-666	TRCN0000007362	USP9X	trt_sh	PC3	96 h	2 µL	-0.5264
8	KDD001_A549_96H:TRCN0000019348:-666	TRCN0000019348	ZNF583	trt_sh	A549	96 h	1 µL	-0.5242
9	CGS001_PC3_144H:PARP4:2	CGS001-143	PARP4	trt_sh.cgs	PC3	144 h	2 µL	-0.5232
10	KDC005_HEPG2_96H:TRCN0000083575:-666	TRCN0000083575	F5	trt_sh	HEPG2	96 h	1.5 µL	-0.5208
11	HOG002_A549_6H:K23	DMSO	DMSO	ctl_vehicle	A549	6 h	0.10%	-0.5191
12	KDD008_HEPG2_96H:TRCN0000009850:-666	TRCN0000009850	MET	trt_sh	HEPG2	96 h	1.5 µL	-0.5159
13	KDB008_HEPG2_96H:TRCN0000010047:-666	TRCN0000010047	RBKS	trt_sh	HEPG2	96 h	2 µL	-0.5149
14	CGS001_SKL_96H:GNPDA1:	CGS001-10007	GNPDA1	trt_sh.cgs	SKL	96 h	10 µL	-0.5146
15	KDC003_A375_96H:TRCN0000019148:-666	TRCN0000019148	CBX7	trt_sh	A375	96 h	1 µL	-0.5132
16	DOS034_A375_24H:BRD-K02515288-001-01-1:5.08	BRD-K02515288	BRD-K02515288	trt_cp	A375	24 h	5 µM	-0.5099
17	BRAF001_A375_24H:BRD-K81144366-003-15-5:0.625	BRD-K81144366	moxisylyte	trt_cp	A375	24 h	500 nM	-0.5095
18	CPC010_MCF7_24H:BRD-K21009077-001-03-3:10	BRD-K21009077	BRD-K21009077	trt_cp	MCF7	24 h	10 µM	-0.5073
19	CSS001_NPC_96H:CTACCAG:1.5	CSS001-CTACCAG	CTACCAG	trt_sh.css	NPC	96 h	1.5 µL	-0.507
20	LJP003_SKBR3_6H:CMAP-L101:100	CMAP-L101	EPG	trt_lig	SKBR3	6 h	100 ng/µL	-0.5064
21	DOSBIO002_A375_24H:BRD-K98945308:10.0074	BRD-K98945308	BRD-K98945308	trt_cp	A375	24 h	10 µM	-0.5026
22	CPC016_SKB_24H:BRD-K49061529-001-01-0:10	BRD-K49061529	BRD-K49061529	trt_cp	SKB	24 h	10 µM	-0.5023
23	KDC002_A375_96H:TRCN0000013207:-666	TRCN0000013207	TEAD2	trt_sh	A375	96 h	1 µL	-0.5016
24	CPC017_ASC_24H:BRD-A95445494-001-02-9:10	BRD-A95445494	maackiain	trt_cp	ASC	24 h	10 µM	-0.5013
25	KDB005_ASC_96H:TRCN0000065236:-666	TRCN0000065236	SEC24B	trt_sh	ASC	96 h	6 µL	-0.5006
26	KDB007_A549_96H:TRCN0000046183:-666	TRCN0000046183	LOXL1	trt_sh	A549	96 h	1.5 µL	-0.4999
27	KDB009_ASC_96H:TRCN0000072213:-666	TRCN0000072213	RFP	ctl_vector	ASC	96 h	6 µL	-0.4994
28	KDA006_PC3_144H:TRCN0000235494:-666	TRCN0000235494	PMS1	trt_sh	PC3	144 h	2 µL	-0.499
29	KDA002_A375_96H:TRCN0000034967:-666	TRCN0000034967	ATIC	trt_sh	A375	96 h	1 µL	-0.4976
30	CPC014_NPC_24H:BRD-K29582115-311-02-0:10	BRD-K29582115	ziprasidone	trt_cp	NPC	24 h	10 µM	-0.4975
31	KDD005_A549_96H:TRCN0000015692:-666	TRCN0000015692	ZNF563	trt_sh	A549	96 h	1 µL	-0.4975
32	CPC013_HT29_6H:BRD-K07395346-001-01-3:10	BRD-K07395346	BRD-K07395346	trt_cp	HT29	6 h	10 µM	-0.4973
33	CPC006_HCC515_24H:BRD-A34817987-001-11-8:10	BRD-A34817987	itraconazole	trt_cp	HCC515	24 h	10 µM	-0.4968
34	KDC001_HCC515_96H:TRCN0000000730:-666	TRCN0000000730	RPS6KB2	trt_sh	HCC515	96 h	3 µL	-0.4962
35	CPD003_PC3_24H:BRD-A99833829-003-19-8:10	BRD-A99833829	bethanechol	trt_cp	PC3	24 h	10 µM	-0.4961
36	ERGK012_VCAP_120H:TRCN0000199080:-666	TRCN0000199080	HIPK3	trt_sh	VCAP	120 h	5 µL	-0.4957
37	KDA010_PC3_144H:TRCN0000050877:-666	TRCN0000050877	DUT	trt_sh	PC3	144 h	2 µL	-0.4956
38	CPC018_SKB_24H:BRD-K41567364-001-01-2:10	BRD-K41567364	SB-334867	trt_cp	SKB	24 h	10 µM	-0.4953
39	MUC.CP002_NKDBA_24H:BRD-K72765491-001-01-0:0.372	BRD-K72765491	SA-1919877	trt_cp	NKDBA	24 h	500 nM	-0.4952
40	CPC010_HA1E_6H:BRD-K20338176-001-14-4:10	BRD-K20338176	cefaclor	trt_cp	HA1E	6 h	10 µM	-0.4946
41	CSS001_PC3_96H:ACACCCT:2	CSS001-ACACCCT	ACACCCT	trt_sh.css	PC3	96 h	2 µL	-0.4946

42	DOS059_HT29_24H:BRD-K61629734-001-01-8:5	BRD-K61629734	BRD-K61629734	trt_cp	HT29	24 h	5 µM	-0.4944
43	CPC019_A549_24H:BRD-K25608435-001-01-9:10	BRD-K25608435	BRD-K25608435	trt_cp	A549	24 h	10 µM	-0.4943
44	KDD003_HA1E_96H:TRCN0000016045:-666	TRCN0000016045	MEIS2	trt_sh	HA1E	96 h	1.5 µL	-0.4943
45	KDC005_A375_96H:TRCN0000010244:-666	TRCN0000010244	ETNK1	trt_sh	A375	96 h	1 µL	-0.4938
46	PCLB003_HCC515_24H:BRD-K64106162-001-02-3:0.04	BRD-K64106162	BRD-K64106162	trt_cp	HCC515	24 h	0.04 µM	-0.4933
47	KDB004_PC3_144H:TRCN0000059922:-666	TRCN0000059922	CASC3	trt_sh	PC3	144 h	2 µL	-0.4927
48	KDB001_SKL_96H:TRCN0000072242:-666	TRCN0000072242	lacZ	ctl_vector	SKL	96 h	10 µL	-0.4925
49	CGS001_HCC515_96H:GCDH:2	CGS001-2639	GCDH	trt_sh.cgs	HCC515	96 h	2 µL	-0.4922
50	DOS008_VCAP_6H:BRD-K82599388-001-01-7:5.03	BRD-K82599388	BRD-K82599388	trt_cp	VCAP	6 h	5 µM	-0.4921
51	CGS001_PC3_144H:NUDT6:2	CGS001-11162	NUDT6	trt_sh.cgs	PC3	144 h	2 µL	-0.4918
52	HSF031_HEK293T_48H:CMAP-HSF-DNMT3A:200	CMAP-HSF-DNMT3A	DNMT3A	trt_oe	HEK293T	48 h	-666	-0.4917
53	KDC008_MCF7_96H:TRCN0000074893:-666	TRCN0000074893	BUD13	trt_sh	MCF7	96 h	2 µL	-0.49
54	KDC003_MCF7_96H:TRCN0000007215:-666	TRCN0000007215	UBE2N	trt_sh	MCF7	96 h	2 µL	-0.4896
55	KDC008_HA1E_96H:TRCN0000014213:-666	TRCN0000014213	VIPR1	trt_sh	HA1E	96 h	1.5 µL	-0.4892
56	KDB005_A375_96H:TRCN0000115682:-666	TRCN0000115682	NCAPD2	trt_sh	A375	96 h	1 µL	-0.4889
57	KDC006_A375_96H:TRCN0000108282:-666	TRCN0000108282	COL4A2	trt_sh	A375	96 h	1 µL	-0.4888
58	KDD006_MCF7_96H:TRCN0000004264:-666	TRCN0000004264	STAT1	trt_sh	MCF7	96 h	2 µL	-0.4888
59	DOS054_A549_24H:BRD-K45429601-001-01-4:5	BRD-K45429601	BRD-K45429601	trt_cp	A549	24 h	5 µM	-0.4883
60	KDB007_A375_96H:TRCN0000046183:-666	TRCN0000046183	LOXL1	trt_sh	A375	96 h	1 µL	-0.4883
61	CYT001_HEPG2_2H:SST:10	SST	SST	trt_lig	HEPG2	2 h	10 ng/mL	-0.4878
62	ERG011_VCAP_24H:BRD-K57080016-001-09-2:0.4	BRD-K57080016	selumetinib	trt_cp	VCAP	24 h	500 nM	-0.4875
63	KDC002_A375_96H:TRCN0000057104:-666	TRCN0000057104	C1QBP	trt_sh	A375	96 h	1 µL	-0.4875
64	DOS039_A549_24H:BRD-K33054146-001-01-6:5.02	BRD-K33054146	BRD-K33054146	trt_cp	A549	24 h	5 µM	-0.4871
65	NMH001_NPC_6H:BRD-K76894938-001-01-5:10	BRD-K76894938	CHIR-99021	trt_cp	NPC	6 h	10 µM	-0.4862
66	CSS001_VCAP_120H:ACATGGT:5	CSS001-ACATGGT	ACATGGT	trt_sh.css	VCAP	120 h	5 µL	-0.4861
67	DOSBIO001_PC3_24H:BRD-K78687850:10.0151	BRD-K78687850	BRD-K78687850	trt_cp	PC3	24 h	10 µM	-0.486
68	CGS001_A375_96H:LOXL1:1	CGS001-4016	LOXL1	trt_sh.cgs	A375	96 h	1 µL	-0.4855
69	KDA005_HEPG2_96H:TRCN0000037582:-666	TRCN0000037582	AURKAIP1	trt_sh	HEPG2	96 h	2 µL	-0.4854
70	KDC006_HEPG2_96H:TRCN0000046191:-666	TRCN0000046191	AOC3	trt_sh	HEPG2	96 h	1.5 µL	-0.4851
71	CPC006_VCAP_24H:A18	DMSO	DMSO	ctl_vehicle	VCAP	24 h	-666	-0.485
72	KDB009_MCF7_144H:TRCN0000049743:-666	TRCN0000049743	PARN	trt_sh	MCF7	144 h	2 µL	-0.4848
73	KDD002_HA1E_96H:TRCN0000019381:-666	TRCN0000019381	BRD8	trt_sh	HA1E	96 h	1.5 µL	-0.4847
74	CPC015_MCF7_24H:BRD-A78391468-001-01-0:10	BRD-A78391468	prednisolone	trt_cp	MCF7	24 h	10 µM	-0.4846
75	CPC019_A549_24H:BRD-K43682718-001-01-0:10	BRD-K43682718	SA-424589	trt_cp	A549	24 h	10 µM	-0.4845
76	KDD009_A549_96H:TRCN0000002832:-666	TRCN0000002832	PTPN12	trt_sh	A549	96 h	1 µL	-0.4845
77	CGS001_A375_96H:RASD1:1	CGS001-51655	RASD1	trt_sh.cgs	A375	96 h	1 µL	-0.4842
78	KDB009_A549_96H:TRCN0000075165:-666	TRCN0000075165	SPEN	trt_sh	A549	96 h	1.5 µL	-0.4842
79	CPC015_SKB_24H:BRD-K49945136-001-03-5:10	BRD-K49945136	GR-113808	trt_cp	SKB	24 h	10 µM	-0.4836
80	KDB008_PC3_144H:TRCN0000062711:-666	TRCN0000327898	BID	trt_sh	PC3	144 h	2 µL	-0.4834
81	KDC001_VCAP_120H:TRCN0000047548:-666	TRCN0000047548	G3BP2	trt_sh	VCAP	120 h	5 µL	-0.4834
82	OEC001_A375_96H:CCSBBROAD304_06290:-666	ccsbBroad304_06290	GNA15	trt_oe	A375	96 h	2 µL	-0.4834
83	CGS001_HCC515_96H:RPA2:2	CGS001-6118	RPA2	trt_sh.cgs	HCC515	96 h	2 µL	-0.4832
84	KDA005_HA1E_96H:TRCN0000036034:-666	TRCN0000036034	OXCT1	trt_sh	HA1E	96 h	2 µL	-0.4831

85	KDD002_HA1E_96H:TRCN0000018964:-666	TRCN0000018964	ZNF354B	trt_sh	HA1E	96 h	1.5 µL	-0.4831
86	CSS001_A549_96H:TGAATCC:1	CSS001-TGAATCC	TGAATCC	trt_sh.css	A549	96 h	1 µL	-0.483
87	KDB009_A375_96H:TRCN0000022296:-666	TRCN0000022296	DNAJB12	trt_sh	A375	96 h	1 µL	-0.4828
88	KDB005_PC3_144H:TRCN0000074719:-666	TRCN0000074719	LSM6	trt_sh	PC3	144 h	2 µL	-0.4827
89	CGS001_HCC515_96H:ZNF114:2	CGS001-163071	ZNF114	trt_sh.cgs	HCC515	96 h	2 µL	-0.4826
90	KDD008_HCC515_96H:TRCN0000007937:-666	TRCN0000007937	PARP2	trt_sh	HCC515	96 h	3 µL	-0.4826
91	KDD010_A549_96H:TRCN0000007937:-666	TRCN0000007937	PARP2	trt_sh	A549	96 h	1 µL	-0.4822
92	KDB007_MCF7_144H:TRCN0000001716:-666	TRCN0000001716	MOK	trt_sh	MCF7	144 h	2 µL	-0.4818
93	CPC009_MCF7_24H:BRD-K28346421-001-01-9:10	BRD-K28346421	rifapentine	trt_cp	MCF7	24 h	10 µM	-0.4817
94	KDD005_MCF7_96H:TRCN0000015689:-666	TRCN0000015689	ZNF563	trt_sh	MCF7	96 h	2 µL	-0.4816
95	KDA001_HA1E_96H:TRCN0000002004:-666	TRCN0000002004	PRKACB	trt_sh	HA1E	96 h	2 µL	-0.4815
96	HOG002_A549_24H:BRD-K12762134-001-01-3:0.005	BRD-K12762134	XAV-939	trt_cp	A549	24 h	1 nM	-0.4812
97	KDD005_A375_96H:TRCN0000365104:-666	TRCN0000365104	ATG4B	trt_sh	A375	96 h	1 µL	-0.4812
98	KDD009_VCAP_120H:TRCN0000040072:-666	TRCN0000040072	BCL2	trt_sh	VCAP	120 h	5 µL	-0.4809
99	PCLB003_A549_24H:BRD-K95309561-001-20-5:1.11	BRD-K95309561	dienestrol	trt_cp	A549	24 h	1.11 µM	-0.4809
100	CGS001_VCAP_120H:RFNG:5	CGS001-5986	RFNG	trt_sh.cgs	VCAP	120 h	5 µL	-0.4807
101	CSS001_A375_96H:ATAGCGC:1	CSS001-ATAGCGC	ATAGCGC	trt_sh.css	A375	96 h	1 µL	-0.4806
102	DOSBIO002_A375_24H:P18	DMSO	DMSO	ctl_vehicle	A375	24 h	0.10%	-0.4806
103	KDD001_A375_96H:TRCN0000047354:-666	TRCN0000047354	RASD1	trt_sh	A375	96 h	1 µL	-0.4805
104	DOS059_HT29_24H:BRD-K88311084-001-01-8:5	BRD-K88311084	BRD-K88311084	trt_cp	HT29	24 h	5 µM	-0.4803
105	CGS001_PC3_96H:LRPPRC:2	CGS001-10128	LRPPRC	trt_sh.cgs	PC3	96 h	2 µL	-0.4802
106	KDB008_ASC_96H:TRCN0000019163:-666	TRCN0000019163	IGHMBP2	trt_sh	ASC	96 h	6 µL	-0.48
107	KDC004_HA1E_96H:TRCN0000062193:-666	TRCN0000062193	SSB	trt_sh	HA1E	96 h	1.5 µL	-0.4799
108	KDC001_VCAP_120H:TRCN0000007937:-666	TRCN0000007937	PARP2	trt_sh	VCAP	120 h	5 µL	-0.4798
109	KDA007_PC3_144H:TRCN0000014176:-666	TRCN0000014176	TBXA2R	trt_sh	PC3	144 h	2 µL	-0.4797
110	KDA007_A549_96H:TRCN0000049362:-666	TRCN0000049362	GNPDA1	trt_sh	A549	96 h	1.5 µL	-0.4794
111	DOS022_VCAP_6H:BRD-K07107626-003-02-7:5.02	BRD-K07107626	BRD-K07107626	trt_cp	VCAP	6 h	5 µM	-0.4793
112	KDC002_HEPG2_96H:TRCN0000013207:-666	TRCN0000013207	TEAD2	trt_sh	HEPG2	96 h	1.5 µL	-0.4791
113	DOS024_VCAP_6H:BRD-K24427624-001-01-6:4.96	BRD-K24427624	BRD-K24427624	trt_cp	VCAP	6 h	5 µM	-0.479
114	OEB002_HEPG2_96H:BRDN0000401093:-666	ccsbBroad304_02092	SYNGR3	trt_oe	HEPG2	96 h	2 µL	-0.479
115	KDC004_HT29_96H:TRCN0000074860:-666	TRCN0000074860	SMU1	trt_sh	HT29	96 h	1 µL	-0.4788
116	KDC010_A375_96H:TRCN0000008715:-666	TRCN0000008715	ATOX1	trt_sh	A375	96 h	1 µL	-0.4786
117	KDC006_HEPG2_96H:TRCN0000064379:-666	TRCN0000064379	GLRX2	trt_sh	HEPG2	96 h	1.5 µL	-0.4783
118	KDD007_A549_96H:TRCN0000074679:-666	TRCN0000074679	LSM2	trt_sh	A549	96 h	1 µL	-0.478
119	TAK002_SW480_96H:TRCN0000014838:-666	TRCN0000014838	ZNF24	trt_sh	SW480	96 h	1 µL	-0.478
120	MUC.CP003_MCF7_24H:A18	DMSO	DMSO	ctl_vehicle	MCF7	24 h	-666	-0.4772
121	KDA005_HT29_96H:TRCN0000355725:-666	TRCN0000355725	MDM2	trt_sh	HT29	96 h	1 µL	-0.4766
122	KDB002_PC3_144H:TRCN0000145055:-666	TRCN0000145055	UTP18	trt_sh	PC3	144 h	2 µL	-0.4766
123	CPC011_VCAP_6H:BRD-A50684349-066-08-3:10	BRD-A50684349	fenoldopam	trt_cp	VCAP	6 h	10 µM	-0.4764
124	CPC019_A549_24H:BRD-K40566034-001-01-5:10	BRD-K40566034	BRD-K40566034	trt_cp	A549	24 h	10 µM	-0.4763
125	CPC010_HCC515_6H:BRD-K70751730-001-03-9:10	BRD-K70751730	BRD-K70751730	trt_cp	HCC515	6 h	10 µM	-0.476
126	CSS001_PC3_144H:ATGCTGT:2	CSS001-ATGCTGT	ATGCTGT	trt_sh.css	PC3	144 h	2 µL	-0.4755
127	OEB005_PC3_96H:BRDN0000410779:-666	ccsbBroad304_02853	QPCT	trt_oe	PC3	96 h	2 µL	-0.4754

128	CPD003_PC3_24H:BRD-K60369935-001-11-8:10	BRD-K60369935	ribavirin	trt_cp	PC3	24 h	10 μM	-0.4748
129	CPC010_HT29_6H:BRD-K15054362-001-04-9:10	BRD-K15054362	BRD-K15054362	trt_cp	HT29	6 h	10 μM	-0.4745
130	CPC016_SKB_24H:BRD-K96402602-001-01-8:10	BRD-K96402602	farnesylthiotriazole	trt_cp	SKB	24 h	10 μM	-0.4744
131	CYT001_PC3_4H:CMAP-CYT-E3641:0.3	CMAP-CYT-E3641	EGFR	trt_lig	PC3	4 h	0.3 ng/mL	-0.4744
132	DOS034_PC3_24H:BRD-K19818442-001-01-3:5.02	BRD-K19818442	BRD-K19818442	trt_cp	PC3	24 h	5 μM	-0.4744
133	CGS001_A375_96H:IPMK:1	CGS001-253430	IPMK	trt_sh.cgs	A375	96 h	1 μL	-0.474
134	CSS001_HT29_96H:GAACGAC:1	CSS001-GAACGAC	GAACGAC	trt_sh.css	HT29	96 h	1 μL	-0.4736
135	LJP001_MCF10A_6H:I18	DMSO	DMSO	ctl_vehicle	MCF10A	6 h	-666	-0.4736
136	KDC003_HCC515_96H:TRCN0000019722:-666	TRCN0000019722	MCM7	trt_sh	HCC515	96 h	3 μL	-0.4734
137	LJP001_MCF10A_24H:BRD-K78431006-001-04-5:0.4	BRD-K78431006	crizotinib	trt_cp	MCF10A	24 h	500 nM	-0.4734
138	PCLB002_HT29_24H:BRD-K98645985:0.12	BRD-K98645985	BRD-K98645985	trt_cp	HT29	24 h	0.12 μM	-0.4734
139	KDD008_A549_96H:TRCN0000000396:-666	TRCN0000000396	MET	trt_sh	A549	96 h	1 μL	-0.4732
140	KDA002_PC3_96H:TRCN0000036119:-666	TRCN0000036119	GSTZ1	trt_sh	PC3	96 h	2 μL	-0.4731
141	KDC008_A549_96H:TRCN0000007480:-666	TRCN0000007480	HSF1	trt_sh	A549	96 h	1 μL	-0.4731
142	KDB007_NPC_96H:TRCN0000036034:-666	TRCN0000036034	OXCT1	trt_sh	NPC	96 h	1.5 μL	-0.4729
143	CPC006_SW480_6H:BRD-A19500257:10	BRD-A19500257	geldanamycin	trt_cp	SW480	6 h	10 μM	-0.4728
144	CGS001_A375_96H:SLC16A2:1	CGS001-6567	SLC16A2	trt_sh.cgs	A375	96 h	1 μL	-0.4727
145	CPC014_PC3_24H:BRD-A93893742-001-07-1:10	BRD-A93893742	EMF-sumo1-2	trt_cp	PC3	24 h	10 μM	-0.4725
146	KDB009_HCC515_96H:TRCN0000008236:-666	TRCN0000008236	CD97	trt_sh	HCC515	96 h	2 μL	-0.4722
147	CPC013_NEU_24H:BRD-U82589721-000-01-4:10	BRD-U82589721	HG-5-113-01	trt_cp	NEU	24 h	10 μM	-0.4719
148	KDB005_A549_96H:TRCN0000019853:-666	TRCN0000019853	ZFP161	trt_sh	A549	96 h	1.5 μL	-0.4718
149	KDD006_VCAP_120H:TRCN0000036629:-666	TRCN0000036629	NDUFS4	trt_sh	VCAP	120 h	5 μL	-0.4718
150	KDB001_A375_96H:TRCN0000166636:-666	TRCN0000166636	IER3	trt_sh	A375	96 h	1 μL	-0.4717
151	TAK002_A375_96H:TRCN0000004790:-666	TRCN0000004790	HECW2	trt_sh	A375	96 h	1 μL	-0.4716
152	CGS001_VCAP_120H:APOM:5	CGS001-55937	APOM	trt_sh.cgs	VCAP	120 h	5 μL	-0.4715
153	KDC001_PC3_96H:TRCN0000034831:-666	TRCN0000034831	METTL4	trt_sh	PC3	96 h	2 μL	-0.4715
154	KDD006_A375_96H:TRCN0000003590:-666	TRCN0000003590	SERPINH1	trt_sh	A375	96 h	1 μL	-0.4714
155	CPC014_A549_6H:BRD-A80574334-001-01-9:10	BRD-A80574334	oxalomalic-acid	trt_cp	A549	6 h	10 μM	-0.4712
156	KDB005_NPC_96H:TRCN0000038636:-666	TRCN0000038636	TCIRG1	trt_sh	NPC	96 h	1.5 μL	-0.471
157	KDA007_PC3_96H:TRCN0000020520:-666	TRCN0000020520	MYBL2	trt_sh	PC3	96 h	2 μL	-0.4708
158	DOS032_PC3_24H:BRD-K24149610-001-01-2:4.91	BRD-K24149610	BRD-K24149610	trt_cp	PC3	24 h	5 μM	-0.4707
159	EKW001_SHSY5Y_120H:TRCN0000295865:2	TRCN0000295865	DDR1	trt_sh	SHSY5Y	120 h	-666	-0.4707
160	KDC010_HEPG2_96H:TRCN0000036793:-666	TRCN0000036793	GPSM1	trt_sh	HEPG2	96 h	1.5 μL	-0.4706
161	CPC006_AGS_6H:BRD-K68997413-001-01-0:10	BRD-K68997413	PF-3845	trt_cp	AGS	6 h	10 μM	-0.4704
162	HOG003_MCF7_24H:I23	DMSO	DMSO	ctl_vehicle	MCF7	24 h	0.10%	-0.4703
163	KDC007_VCAP_120H:TRCN0000053444:-666	TRCN0000053444	SMOC2	trt_sh	VCAP	120 h	5 μL	-0.4702
164	KDD001_PC3_96H:TRCN0000019348:-666	TRCN0000019348	ZNF583	trt_sh	PC3	96 h	2 μL	-0.4701
165	DOSBIO001_MCF7_24H:BRD-K36928794:10.2341	BRD-K36928794	BRD-K36928794	trt_cp	MCF7	24 h	10 μM	-0.47
166	ERGK004_VCAP_120H:TRCN0000037476:-666	TRCN0000037476	FLJ40852	trt_sh	VCAP	120 h	5 μL	-0.4699
167	DOSBIO001_A375_24H:B06	DMSO	DMSO	ctl_vehicle	A375	24 h	0.10%	-0.4696
168	KDB005_A375_96H:TRCN0000013229:-666	TRCN0000013229	CSDA	trt_sh	A375	96 h	1 μL	-0.4696
169	CPC019_HA1E_6H:BRD-A32595718-001-01-9:10	BRD-A32595718	BJM-CSC-15	trt_cp	HA1E	6 h	10 μM	-0.4695
170	KDD002_HT29_96H:TRCN0000002860:-666	TRCN0000002860	PTPRG	trt_sh	HT29	96 h	1 μL	-0.4694

171	KDD001_A549_96H:TRCN0000072419:-666	TRCN0000072419	LMNB2	trt_sh	A549	96 h	1 µL	-0.4692
172	OEC001_A375_72H:CCSBBROAD304_02030:-666	ccsbBroad304_02030	CFLAR	trt_oe	A375	72 h	2 µL	-0.4689
173	KDD002_A549_96H:TRCN0000006788:-666	TRCN0000006788	F10	trt_sh	A549	96 h	1 µL	-0.4688
174	PCLB002_MCF7_24H:BRD-K39569857:0.12	BRD-K39569857	avrainvillamide-analog-3	trt_cp	MCF7	24 h	0.12 µM	-0.4688
175	CPC018_MCF7_24H:BRD-K20313525-001-03-8:10	BRD-K20313525	rosmarinic-acid	trt_cp	MCF7	24 h	10 µM	-0.4687
176	KDD008_HEPG2_96H:TRCN0000039947:-666	TRCN0000039947	CHEK2	trt_sh	HEPG2	96 h	1.5 µL	-0.4687
177	KDD002_A549_96H:TRCN0000059100:-666	TRCN0000059100	IFNGR2	trt_sh	A549	96 h	1 µL	-0.4686
178	KDA004_PC3_96H:TRCN0000058715:-666	TRCN0000058715	TNFRSF1A	trt_sh	PC3	96 h	2 µL	-0.4685
179	OEB001_PC3_96H:BRDN0000399999:-666	ccsbBroad304_02889	WWTR1	trt_oe	PC3	96 h	2 µL	-0.4685
180	CSS001_A375_96H:AAATCCA:1	CSS001-AAATCCA	AAATCCA	trt_sh.css	A375	96 h	1 µL	-0.4681
181	DOS015_VCAP_24H:BRD-K19203541-001-01-8:5.01	BRD-K19203541	BRD-K19203541	trt_cp	VCAP	24 h	5 µM	-0.4681
182	CPC010_HEPG2_6H:BRD-K39339537-001-18-6:10	BRD-K39339537	epirizole	trt_cp	HEPG2	6 h	10 µM	-0.468
183	CPC015_PHH_24H:BRD-K21733600-001-06-7:10	BRD-K21733600	rofecoxib	trt_cp	PHH	24 h	10 µM	-0.468
184	CPC020_PC3_24H:BRD-A25067867-066-14-3:10	BRD-A25067867	benzatropine	trt_cp	PC3	24 h	10 µM	-0.468
185	CPC001_VCAP_6H:BRD-K29178788-001-02-0:10	BRD-K29178788	dictamnine	trt_cp	VCAP	6 h	10 µM	-0.4679
186	KDA002_A375_96H:TRCN0000035049:-666	TRCN0000035049	HPRT1	trt_sh	A375	96 h	1 µL	-0.4675
187	DOS015_VCAP_6H:BRD-K59330510-001-01-7:5.03	BRD-K59330510	BRD-K59330510	trt_cp	VCAP	6 h	5 µM	-0.4674
188	ERG014_VCAP_24H:BRD-K40690989-001-01-8:10.0451	BRD-K40690989	BRD-K40690989	trt_cp	VCAP	24 h	10 µM	-0.4673
189	KDB009_HA1E_96H:TRCN0000039935:-666	TRCN0000039935	PPP2R5B	trt_sh	HA1E	96 h	2 µL	-0.4673
190	OEB006_HCC515_96H:BRDN0000461475:-666	ccsbBroad304_07064	TH	trt_oe	HCC515	96 h	4 µL	-0.4671
191	TAK002_SW480_96H:TRCN0000431687:-666	TRCN0000431687	CD1C	trt_sh	SW480	96 h	1 µL	-0.4671
192	CPC013_ASC_24H:BRD-K89784888-001-01-3:10	BRD-K89784888	BRD-K89784888	trt_cp	ASC	24 h	10 µM	-0.467
193	CPC006_NCIH1836_6H:BRD-K80527266-001-01-7:10	BRD-K80527266	triacsin-c	trt_cp	NCIH1836	6 h	10 µM	-0.4667
194	CPC014_VCAP_6H:M10	DMSO	DMSO	ctl_vehicle	VCAP	6 h	-666	-0.4667
195	CPC014_PC3_24H:BRD-A80788753-001-01-7:10	BRD-A80788753	EMF-sumo1-10	trt_cp	PC3	24 h	10 µM	-0.4666
196	CPC015_HEPG2_6H:B17	DMSO	DMSO	ctl_vehicle	HEPG2	6 h	-666	-0.4666
197	CPC019_MCF7_24H:BRD-K68690198-001-01-2:10	BRD-K68690198	BRD-K68690198	trt_cp	MCF7	24 h	10 µM	-0.4665
198	CPC013_PC3_24H:BRD-A38275906-001-01-4:10	BRD-A38275906	ST-019366	trt_cp	PC3	24 h	10 µM	-0.4664
199	ERG010_VCAP_24H:BRD-K44916273-001-01-8:9.321	BRD-K44916273	BRD-K44916273	trt_cp	VCAP	24 h	10 µM	-0.4664
200	KDB001_A375_96H:TRCN0000285376:-666	TRCN0000285376	SHB	trt_sh	A375	96 h	1 µL	-0.4663

Legend: the ranked highest connectivity score for reversing the effect is presented with identification of the signature in the database (sig_id), identification of the chemical or genetic manipulation (pert_id), frequent name of the perturbagen (pert_iname), type of perturbation (pert_type), cell type (cell_id), duration (pert_itime), and dose of treatment (pert_idose) and the respective connectivity score (score). More information can be obtained on http://apps.lincscloud.org.

Appendix 3B. Connectivity mapping results for top 200 compounds mimicking the *NT5C2* knockdown signature in all cell lines available on LincsCloud.

Rank	MIMICKING SIGNATURE sig_id	pert_id	pert_iname	pert_type	cell_id	pert_itime	pert_idose	score
1	CPC008_HEPG2_6H:BRD-K33720394-001-01-1:10	BRD-K33720394	BRD-K33720394	trt_cp	HEPG2	6 h	10 µM	0.5993
2	DOS024_VCAP_24H:H20	DMSO	DMSO	ctl_vehicle	VCAP	24 h	-666	0.5952
3	KDD002_HA1E_96H:TRCN0000063438:-666	TRCN0000063438	GPR153	trt_sh	HA1E	96 h	1.5 µL	0.5717
4	KDC005_A549_96H:TRCN0000018557:-666	TRCN0000018557	NFKBIB	trt_sh	A549	96 h	1 µL	0.5611
5	DOS015_VCAP_24H:BRD-K62039221-001-01-9:5.03	BRD-K62039221	BRD-K62039221	trt_cp	VCAP	24 h	5 µM	0.557
6	DOS023_VCAP_24H:BRD-K86843304-001-02-4:5	BRD-K86843304	BRD-K86843304	trt_cp	VCAP	24 h	5 µM	0.5567
7	KDB002_PC3_96H:TRCN0000330960:-666	TRCN0000330960	UTP14A	trt_sh	PC3	96 h	2 µL	0.5546
8	KDD009_A549_96H:TRCN0000010295:-666	TRCN0000010295	IGBP1	trt_sh	A549	96 h	1 µL	0.5468
9	KDA001_A549_96H:TRCN0000061756:-666	TRCN0000061756	CAT	trt_sh	A549	96 h	1.5 µL	0.5458
10	DOS021_VCAP_24H:BRD-K35421408-001-02-6:5.01	BRD-K35421408	BRD-K35421408	trt_cp	VCAP	24 h	5 µM	0.545
11	OEB002_PC3_96H:BRDN0000401187:-666	ccsbBroad304_01579	SOX2	trt_oe	PC3	96 h	2 µL	0.5442
12	DOS027_VCAP_24H:BRD-K15958865-001-01-9:5.02	BRD-K15958865	BRD-K15958865	trt_cp	VCAP	24 h	5 µM	0.5438
13	KDB006_PC3_96H:TRCN0000036550:-666	TRCN0000036550	MTHFD2	trt_sh	PC3	96 h	2 µL	0.5434
14	CPC015_A549_24H:BRD-A25143711-001-01-4:10	BRD-A25143711	hydrocortisone	trt_cp	A549	24 h	10 µM	0.5425
15	DER001_A375_96H:TRCN0000049185:-666	TRCN0000049185	MPI	trt_sh	A375	96 h	1 µL	0.5414
16	KDB004_PC3_144H:TRCN0000062351:-666	TRCN0000062351	GADD45A	trt_sh	PC3	144 h	2 µL	0.5401
17	DOS023_VCAP_24H:BRD-K56553629-001-01-5:5.03	BRD-K56553629	BRD-K56553629	trt_cp	VCAP	24 h	5 µM	0.5375
18	DOS027_VCAP_24H:BRD-K50727748-001-01-7:5.01	BRD-K50727748	BRD-K50727748	trt_cp	VCAP	24 h	5 µM	0.5345
19	DOS022_VCAP_24H:BRD-K45319854-001-01-1:5.04	BRD-K45319854	BRD-K45319854	trt_cp	VCAP	24 h	5 µM	0.5325
20	CGS001_A375_96H:NPTN:1	CGS001-27020	NPTN	trt_sh.cgs	A375	96 h	1 µL	0.5322
21	KDB005_PC3_96H:TRCN0000028456:-666	TRCN0000028456	BCKDHA	trt_sh	PC3	96 h	2 µL	0.5284
22	KDA007_HCC515_96H:TRCN0000327648:-666	TRCN0000327648	FBXL20	trt_sh	HCC515	96 h	2 µL	0.528
23	DOS020_VCAP_24H:BRD-K81156227-001-01-2:4.97	BRD-K81156227	BRD-K81156227	trt_cp	VCAP	24 h	5 µM	0.5272
24	CGS001_A549_96H:RBP4:1	CGS001-5950	RBP4	trt_sh.cgs	A549	96 h	1 µL	0.5267
25	KDB005_PC3_96H:TRCN0000072194:-666	TRCN0000072194	GFP	ctl_vector	PC3	96 h	2 µL	0.5248
26	CYT001_HT29_2H:FGF18:20	FGF18	FGF18	trt_lig	HT29	2 h	20 ng/mL	0.5245
27	OEB006_HCC515_96H:BRDN0000461135:-666	ccsbBroad304_13929	PRKACA	trt_oe	HCC515	96 h	4 µL	0.5244
28	CPC005_A549_6H:BRD-A37780065-001-03-7:10	BRD-A37780065	triamcinolone	trt_cp	A549	6 h	10 µM	0.5237
29	ERGK008_VCAP_120H:TRCN0000010221:-666	TRCN0000010221	NME7	trt_sh	VCAP	120 h	5 µL	0.5229
30	DOS024_VCAP_24H:BRD-K20051162-001-01-6:5.05	BRD-K20051162	BRD-K20051162	trt_cp	VCAP	24 h	5 µM	0.5228
31	DOS023_VCAP_24H:BRD-K12907790-001-01-9:4.92	BRD-K12907790	BRD-K12907790	trt_cp	VCAP	24 h	5 µM	0.5218
32	DOS018_VCAP_24H:BRD-K07523222-001-01-9:5	BRD-K07523222	BRD-K07523222	trt_cp	VCAP	24 h	5 µM	0.5217
33	KDC010_A549_96H:TRCN0000003109:-666	TRCN0000003109	MST1R	trt_sh	A549	96 h	1 µL	0.5215
34	CPC006_JHUEM2_6H:BRD-K53903639-001-01-3:80	BRD-K53903639	CHEMBL-1222381	trt_cp	JHUEM2	6 h	80 µM	0.5212
35	KDB009_MCF7_144H:TRCN0000014714:-666	TRCN0000014714	RELB	trt_sh	MCF7	144 h	2 µL	0.5206
36	CSS001_VCAP_120H:AGCTGAC:5	CSS001-AGCTGAC	AGCTGAC	trt_sh.css	VCAP	120 h	5 µL	0.5192
37	KDA008_PC3_144H:TRCN0000008673:-666	TRCN0000008673	ZNF300	trt_sh	PC3	144 h	2 µL	0.5186
38	KDC009_VCAP_120H:TRCN0000033612:-666	TRCN0000033612	BBC3	trt_sh	VCAP	120 h	5 µL	0.5185
39	DOS018_VCAP_24H:BRD-K89621861-001-01-8:4.94	BRD-K89621861	BRD-K89621861	trt_cp	VCAP	24 h	5 µM	0.5181
40	CGS001_A549_96H:LHX8:1	CGS001-431707	LHX8	trt_sh.cgs	A549	96 h	1 µL	0.518
41	KDC008_VCAP_120H:TRCN0000013733:-666	TRCN0000013733	NKX2-5	trt_sh	VCAP	120 h	5 µL	0.5178

42	KDB009_HEPG2_96H:TRCN0000008876:-666	TRCN0000008876	GPER	trt_sh	HEPG2	96 h	2 µL	0.5177
43	CPC010_PC3_24H:BRD-K51155404-019-03-3:10	BRD-K51155404	BRD-K51155404	trt_cp	PC3	24 h	10 µM	0.5175
44	KDB002_PC3_96H:TRCN0000233302:-666	TRCN0000233302	TIMM17B	trt_sh	PC3	96 h	2 µL	0.5166
45	KDB006_PC3_96H:TRCN0000039797:-666	TRCN0000039797	AKT1	trt_sh	PC3	96 h	2 µL	0.5161
46	CGS001_A375_96H:CSNK1A1:1	CGS001-1452	CSNK1A1	trt_sh.cgs	A375	96 h	1 µL	0.5159
47	KDB005_PC3_96H:TRCN0000065159:-666	TRCN0000065159	SEC24C	trt_sh	PC3	96 h	2 µL	0.5147
48	KDB006_PC3_96H:TRCN0000043650:-666	TRCN0000043650	SLC2A6	trt_sh	PC3	96 h	2 µL	0.5132
49	KDC004_HEPG2_96H:TRCN0000058890:-666	TRCN0000058890	TNFRSF13C	trt_sh	HEPG2	96 h	1.5 µL	0.5132
50	KDD008_A549_96H:TRCN0000040064:-666	TRCN0000040064	WT1	trt_sh	A549	96 h	1 µL	0.5128
51	CGS001_HCC515_96H:ORC1:2	CGS001-4998	ORC1	trt_sh.cgs	HCC515	96 h	2 µL	0.5122
52	KDC008_A375_96H:TRCN0000062910:-666	TRCN0000062910	PVR	trt_sh	A375	96 h	1 µL	0.5111
53	CPC007_A549_6H:BRD-K59761766-001-05-5:10	BRD-K59761766	BRD-K59761766	trt_cp	A549	6 h	10 µM	0.5109
54	DOS009_VCAP_24H:BRD-K27571934-001-01-3:5.02	BRD-K27571934	BRD-K27571934	trt_cp	VCAP	24 h	5 µM	0.5109
55	CGS001_PC3_96H:PRKCH:2	CGS001-5583	PRKCH	trt_sh.cgs	PC3	96 h	2 µL	0.5104
56	DOS012_VCAP_24H:BRD-K39412479-001-01-7:4.92	BRD-K39412479	BRD-K39412479	trt_cp	VCAP	24 h	5 µM	0.5104
57	CGS001_MCF7_144H:P2RY2:2	CGS001-5029	P2RY2	trt_sh.cgs	MCF7	144 h	2 µL	0.5101
58	TAK001_PC3_96H:TRCN0000047710:-666	TRCN0000047710	RHOA	trt_sh	PC3	96 h	2 µL	0.5094
59	ERGK003_VCAP_120H:TRCN0000199619:-666	TRCN0000199619	KSR2	trt_sh	VCAP	120 h	5 µL	0.5092
60	KDB010_A375_96H:TRCN0000331232:-666	TRCN0000331232	GNAS	trt_sh	A375	96 h	1 µL	0.5091
61	DOS006_VCAP_24H:BRD-K01008530-001-01-4:5.01	BRD-K01008530	BRD-K01008530	trt_cp	VCAP	24 h	5 µM	0.5083
62	KDB005_PC3_96H:TRCN0000050778:-666	TRCN0000050778	ERCC5	trt_sh	PC3	96 h	2 µL	0.5083
63	CPC016_HT29_6H:BRD-A64977602-001-04-3:10	BRD-A64977602	mirtazapine	trt_cp	HT29	6 h	10 µM	0.5079
64	KDB008_MCF7_144H:TRCN0000065023:-666	TRCN0000065023	STX4	trt_sh	MCF7	144 h	2 µL	0.5079
65	OEB001_HA1E_96H:BRDN0000399624:-666	ccsbBroad304_02855	SPDEF	trt_oe	HA1E	96 h	1 µL	0.5078
66	KDC009_HCC515_96H:TRCN0000013745:-666	TRCN0000013745	CTBP2	trt_sh	HCC515	96 h	3 µL	0.5077
67	KDC008_A549_96H:TRCN0000039900:-666	TRCN0000039900	ABL1	trt_sh	A549	96 h	1 µL	0.5075
68	KDC007_VCAP_120H:TRCN0000064730:-666	TRCN0000064730	ME1	trt_sh	VCAP	120 h	5 µL	0.5072
69	CPC009_A549_24H:BRD-K18724229-019-01-5:10	BRD-K18724229	BRD-K18724229	trt_cp	A549	24 h	10 µM	0.5066
70	CSS001_VCAP_120H:AGCTACT:5	CSS001-AGCTACT	AGCTACT	trt_sh.css	VCAP	120 h	5 µL	0.5066
71	KDC007_A375_96H:TRCN0000074712:-666	TRCN0000074712	RPS3A	trt_sh	A375	96 h	1 µL	0.5066
72	KDB002_PC3_96H:TRCN0000135483:-666	TRCN0000135483	TMEM97	trt_sh	PC3	96 h	2 µL	0.5065
73	ERGK008_VCAP_120H:TRCN0000037807:-666	TRCN0000037807	PRKAR2A	trt_sh	VCAP	120 h	5 µL	0.5063
74	DOS027_VCAP_24H:BRD-M66254575-001-01-0:4.99	BRD-M66254575	BRD-M66254575	trt_cp	VCAP	24 h	5 µM	0.5061
75	CPC015_A549_24H:BRD-A02180903-001-03-7:10	BRD-A02180903	betamethasone	trt_cp	A549	24 h	10 µM	0.5054
76	DOS030_VCAP_24H:BRD-K00547540-001-02-8:5.01	BRD-K00547540	BRD-K00547540	trt_cp	VCAP	24 h	5 µM	0.5053
77	KDC006_A549_96H:TRCN0000000478:-666	TRCN0000000478	CAMK2G	trt_sh	A549	96 h	1 µL	0.5048
78	KDD009_HT29_96H:TRCN0000082449:-666	TRCN0000082449	C9ORF96	trt_sh	HT29	96 h	1 µL	0.5046
79	DOS022_VCAP_24H:BRD-K85630885-001-01-5:5.06	BRD-K85630885	BRD-K85630885	trt_cp	VCAP	24 h	5 µM	0.5044
80	CSS001_A549_96H:AAGACAA:1	CSS001-AAGACAA	AAGACAA	trt_sh.css	A549	96 h	1 µL	0.5043
81	KDC003_A549_96H:TRCN0000057941:-666	TRCN0000057941	CXCL1	trt_sh	A549	96 h	1 µL	0.5043
82	DOS053_A549_24H:BRD-K68166950-001-01-1:4.98	BRD-K68166950	BRD-K68166950	trt_cp	A549	24 h	5 µM	0.504
83	KDA002_MCF7_96H:TRCN0000019981:-666	TRCN0000019981	TFE3	trt_sh	MCF7	96 h	2 µL	0.5031
84	DOS024_VCAP_24H:N09	DMSO	DMSO	ctl_vehicle	VCAP	24 h	-666	0.5027

85	CSS001_HA1E_96H:ACAGCAT:1.5	CSS001-ACAGCAT	ACAGCAT	trt_sh.css	HA1E	96 h	1.5 µL	0.5021
86	DOS023_VCAP_24H:BRD-K35493431-001-01-8:5.06	BRD-K35493431	BRD-K35493431	trt_cp	VCAP	24 h	5 µM	0.502
87	CPC006_PL21_6H:BRD-K70792160-003-02-0:24	BRD-K70792160	10-DEBC	trt_cp	PL21	6 h	20 µM	0.5019
88	CPC007_A375_24H:BRD-K74029109-001-05-4:10	BRD-K74029109	BRD-K74029109	trt_cp	A375	24 h	10 µM	0.5008
89	KDA007_A375_96H:TRCN0000040217:-666	TRCN0000040217	IGF1	trt_sh	A375	96 h	1 µL	0.5005
90	KDA006_A375_96H:TRCN0000072249:-666	TRCN0000072249	LUCIFERASE	ctl_vector	A375	96 h	1 µL	0.5002
91	CPC019_PC3_24H:BRD-K60067222-001-01-1:10	BRD-K60067222	BRD-K60067222	trt_cp	PC3	24 h	10 µM	0.5001
92	KDB006_VCAP_120H:TRCN0000056269:-666	TRCN0000056269	PLSCR1	trt_sh	VCAP	120 h	6 µL	0.4996
93	CPC003_VCAP_6H:BRD-A73909368-001-01-5:10	BRD-A73909368	dactinomycin	trt_cp	VCAP	6 h	10 µM	0.4995
94	OEB006_A549_96H:BRDN0000411845:-666	ccsbBroad304_03264	UBAP1	trt_oe	A549	96 h	2 µL	0.499
95	CSS001_A375_96H:ACATCTC:1	CSS001-ACATCTC	ACATCTC	trt_sh.css	A375	96 h	1 µL	0.4989
96	KDA004_MCF7_144H:TRCN0000036012:-666	TRCN0000036012	CROT	trt_sh	MCF7	144 h	2 µL	0.4987
97	KDC010_A549_96H:TRCN0000022042:-666	TRCN0000022042	PHF21B	trt_sh	A549	96 h	1 µL	0.4985
98	CPD001_MCF7_6H:BRD-K67783091-001-20-4:10	BRD-K67783091	haloperidol	trt_cp	MCF7	6 h	10 µM	0.4983
99	DOS027_VCAP_24H:BRD-K02861439-001-01-7:5	BRD-K02861439	BRD-K02861439	trt_cp	VCAP	24 h	5 µM	0.4981
100	KDB005_PC3_96H:TRCN0000055838:-666	TRCN0000055838	CETN3	trt_sh	PC3	96 h	2 µL	0.4979
101	KDC010_MCF7_96H:TRCN0000059049:-666	TRCN0000059049	MTSS1	trt_sh	MCF7	96 h	2 µL	0.4979
102	PCLB002_MCF7_24H:BRD-K81418486:3.33	BRD-K81418486	vorinostat	trt_cp	MCF7	24 h	3.33 µM	0.4978
103	KDC010_A375_96H:TRCN0000053022:-666	TRCN0000053022	PAPOLA	trt_sh	A375	96 h	1 µL	0.4973
104	KDB006_HT29_96H:TRCN0000036550:-666	TRCN0000036550	MTHFD2	trt_sh	HT29	96 h	1 µL	0.4972
105	KDD005_HA1E_96H:TRCN0000018143:-666	TRCN0000018143	ZNF554	trt_sh	HA1E	96 h	1.5 µL	0.4972
106	KDA005_MCF7_144H:TRCN0000039837:-666	TRCN0000039837	BRCA1	trt_sh	MCF7	144 h	2 µL	0.4971
107	DOS043_A549_24H:BRD-K07796595-001-01-0:5.02	BRD-K07796595	BRD-K07796595	trt_cp	A549	24 h	5 µM	0.4965
108	KDC006_VCAP_120H:TRCN0000029204:-666	TRCN0000029204	PHB	trt_sh	VCAP	120 h	5 µL	0.4963
109	KDC005_A375_96H:TRCN0000008147:-666	TRCN0000008147	TMEM11	trt_sh	A375	96 h	1 µL	0.496
110	KDB009_PC3_96H:TRCN0000043019:-666	TRCN0000043019	ATP6V1D	trt_sh	PC3	96 h	2 µL	0.4958
111	CSS001_A549_96H:TCTCGGC:1	CSS001-TCTCGGC	TCTCGGC	trt_sh.css	A549	96 h	1 µL	0.4955
112	CSS001_VCAP_120H:TCATCTG:5	CSS001-TCATCTG	TCATCTG	trt_sh.css	VCAP	120 h	5 µL	0.4955
113	CPC014_NPC_24H:BRD-A75931230-001-01-3:10	BRD-A75931230	F-1061-0166	trt_cp	NPC	24 h	10 µM	0.4953
114	KDD009_A549_96H:TRCN0000040076:-666	TRCN0000040076	EZH2	trt_sh	A549	96 h	1 µL	0.4953
115	DER001_PC3_96H:TRCN0000072249:-666	TRCN0000072249	LUCIFERASE	ctl_vector	PC3	96 h	-666	0.4951
116	KDA003_MCF7_144H:TRCN0000009482:-666	TRCN0000009482	P2RY2	trt_sh	MCF7	144 h	2 µL	0.495
117	CGS001_A375_96H:RPS3A:1	CGS001-6189	RPS3A	trt_sh.cgs	A375	96 h	1 µL	0.4946
118	CGS001_PC3_96H:EPB41L4B:2	CGS001-54566	EPB41L4B	trt_sh.cgs	PC3	96 h	2 µL	0.4945
119	KDD003_A549_96H:TRCN0000021046:-666	TRCN0000021046	ASCL4	trt_sh	A549	96 h	1 µL	0.4944
120	ERGK005_VCAP_120H:TRCN0000037785:-666	TRCN0000037785	PGK2	trt_sh	VCAP	120 h	5 µL	0.4943
121	CGS001_A549_96H:BHMT2:1	CGS001-23743	BHMT2	trt_sh.cgs	A549	96 h	1 µL	0.4941
122	DOS024_VCAP_24H:BRD-K43279914-001-01-7:5.07	BRD-K43279914	BRD-K43279914	trt_cp	VCAP	24 h	5 µM	0.494
123	CPC016_NEU_24H:BRD-K85015012-003-01-1:10	BRD-K85015012	NNC-05-2090	trt_cp	NEU	24 h	10 µM	0.4939
124	CPC014_PC3_24H:BRD-A67748489-001-06-8:10	BRD-A67748489	K3644	trt_cp	PC3	24 h	10 µM	0.4938
125	KDD002_A375_96H:TRCN0000017567:-666	TRCN0000017567	ZEB1	trt_sh	A375	96 h	1 µL	0.4936
126	CGS001_A375_96H:KCNK15:1	CGS001-60598	KCNK15	trt_sh.cgs	A375	96 h	1 µL	0.4935
127	DER001_HT29_96H:TRCN0000006170:-666	TRCN0000006170	GNE	trt_sh	HT29	96 h	1 µL	0.4935

128	KDD007_A375_96H:TRCN0000008095:-666	TRCN0000008095	AGTR1	trt_sh	A375	96 h	1 µL	0.4934
129	LJP001_MCF10A_24H:BRD-K09186807-001-02-4:10	BRD-K09186807	KIN001-244	trt_cp	MCF10A	24 h	10 µM	0.4933
130	KDB004_PC3_96H:TRCN0000019362:-666	TRCN0000019362	KAT6B	trt_sh	PC3	96 h	2 µL	0.4931
131	CPC005_HT29_6H:BRD-K51476772-001-01-9:10	BRD-K51476772	ST-638	trt_cp	HT29	6 h	10 µM	0.4929
132	KDC006_HT29_96H:TRCN0000056886:-666	TRCN0000056886	TRAF1	trt_sh	HT29	96 h	1 µL	0.4929
133	CPC001_VCAP_24H:BRD-K59650319-003-01-7:10	BRD-K59650319	YM-298198	trt_cp	VCAP	24 h	10 µM	0.4926
134	CSS001_HA1E_96H:TCAACGA:1.5	CSS001-TCAACGA	TCAACGA	trt_sh.css	HA1E	96 h	1.5 µL	0.4925
135	KDD007_A375_96H:TRCN0000020543:-666	TRCN0000020543	TWIST1	trt_sh	A375	96 h	1 µL	0.4924
136	DOS023_VCAP_24H:G19	DMSO	DMSO	ctl_vehicle	VCAP	24 h	-666	0.4922
137	DOS023_VCAP_24H:BRD-K43438744-001-02-3:5	BRD-K43438744	BRD-K43438744	trt_cp	VCAP	24 h	5 µM	0.4921
138	OEB003_A375_96H:BRDN0000403688:-666	ccsbBroad304_01240	POLR2F	trt_oe	A375	96 h	2 µL	0.492
139	OEB006_HCC515_96H:BRDN0000462046:-666	ccsbBroad304_00174	BRAF	trt_oe	HCC515	96 h	4 µL	0.492
140	CPC014_NPC_24H:BRD-K66902379-001-01-4:10	BRD-K66902379	BRD-K66902379	trt_cp	NPC	24 h	10 µM	0.4917
141	CPC006_SKM1_6H:BRD-K62810658-001-06-4:10	BRD-K62810658	PD-98059	trt_cp	SKM1	6 h	10 µM	0.4915
142	KDC006_A375_96H:TRCN0000058383:-666	TRCN0000058383	IL1B	trt_sh	A375	96 h	1 µL	0.4915
143	KDC003_A549_96H:TRCN0000008623:-666	TRCN0000008623	CDX1	trt_sh	A549	96 h	1 µL	0.491
144	CPC006_HCC515_24H:BRD-K25737009-001-02-0:40	BRD-K25737009	CD-1530	trt_cp	HCC515	24 h	40 µM	0.4909
145	CSS001_A549_96H:TTAGATG:1	CSS001-TTAGATG	TTAGATG	trt_sh.css	A549	96 h	1 µL	0.4906
146	CPC015_NPC_24H:BRD-A59174698-003-12-8:10	BRD-A59174698	ritodrine	trt_cp	NPC	24 h	10 µM	0.4903
147	KDA002_A375_96H:TRCN0000001694:-666	TRCN0000001694	PRKCA	trt_sh	A375	96 h	1 µL	0.4903
148	KDB003_MCF7_144H:TRCN0000136175:-666	TRCN0000136175	DOK4	trt_sh	MCF7	144 h	2 µL	0.4901
149	DOS006_VCAP_24H:BRD-K32294814-001-02-8:5.01	BRD-K32294814	BRD-K32294814	trt_cp	VCAP	24 h	5 µM	0.49
150	KDB008_VCAP_120H:TRCN0000059370:-666	TRCN0000059370	TM9SF3	trt_sh	VCAP	120 h	6 µL	0.4899
151	KDC001_A549_96H:TRCN0000003371:-666	TRCN0000003371	UBE3A	trt_sh	A549	96 h	1 µL	0.4899
152	KDB007_A375_96H:TRCN0000052664:-666	TRCN0000052664	PCK2	trt_sh	A375	96 h	1 µL	0.4898
153	KDC002_A549_96H:TRCN0000078434:-666	TRCN0000078434	ODC1	trt_sh	A549	96 h	1 µL	0.4897
154	CPC007_HA1E_6H:BRD-A47816767-001-11-1:10	BRD-A47816767	BRD-A47816767	trt_cp	HA1E	6 h	10 µM	0.4896
155	CPC006_PC3_6H:BRD-A52650764-001-01-0:10	BRD-A52650764	ingenol	trt_cp	PC3	6 h	10 µM	0.4894
156	KDD001_A549_96H:TRCN0000036516:-666	TRCN0000036516	ARPC5	trt_sh	A549	96 h	1 µL	0.4894
157	CGS001_A549_96H:DNTTIP2:1	CGS001-30836	DNTTIP2	trt_sh.cgs	A549	96 h	1 µL	0.4892
158	CPC014_A549_24H:BRD-K53972329-001-02-1:10	BRD-K53972329	ruxolitinib	trt_cp	A549	24 h	10 µM	0.4892
159	KDA008_PC3_144H:TRCN0000318417:-666	TRCN0000318417	PIN1	trt_sh	PC3	144 h	2 µL	0.4892
160	CPC007_HCC515_6H:BRD-K72864428-001-02-7:10	BRD-K72864428	BRD-K72864428	trt_cp	HCC515	6 h	10 µM	0.4891
161	CPC013_NEU_24H:BRD-K47731572-001-01-6:10	BRD-K47731572	BRD-K47731572	trt_cp	NEU	24 h	10 µM	0.4888
162	LJP001_MDAMB231_24H:BRD-K24515980-001-01-6:10	BRD-K24515980	QL-XI-92	trt_cp	MDAMB231	24 h	10 µM	0.4887
163	CGS001_HT29_96H:TUBD1:1	CGS001-51174	TUBD1	trt_sh.cgs	HT29	96 h	1 µL	0.4885
164	CSS001_PC3_96H:TGGATCA:2	CSS001-TGGATCA	TGGATCA	trt_sh.css	PC3	96 h	2 µL	0.4884
165	CGS001_A375_96H:TPR:1	CGS001-7175	TPR	trt_sh.cgs	A375	96 h	1 µL	0.488
166	KDC002_A549_96H:TRCN0000018985:-666	TRCN0000018985	SOX10	trt_sh	A549	96 h	1 µL	0.4879
167	KDD001_A549_96H:TRCN0000043907:-666	TRCN0000043907	CLCN5	trt_sh	A549	96 h	1 µL	0.4879
168	KDA004_A375_96H:TRCN0000046324:-666	TRCN0000046324	COX7B	trt_sh	A375	96 h	1 µL	0.4878
169	CPC006_SKLU1_6H:BRD-K19894101-001-01-6:11.1	BRD-K19894101	MST-312	trt_cp	SKLU1	6 h	10 µM	0.4877
170	KDC007_HEPG2_96H:TRCN0000015529:-666	TRCN0000015529	OTX1	trt_sh	HEPG2	96 h	1.5 µL	0.4877

171	CPC008_A375_24H:BRD-K30351863-019-01-2:10	BRD-K30351863	BRD-K30351863	trt_cp	A375	24 h	10 µM	0.4876
172	KDB008_HEPG2_96H:TRCN0000073920:-666	TRCN0000073920	PEPD	trt_sh	HEPG2	96 h	2 µL	0.4876
173	CSS001_A375_96H:TCTGGGA:1	CSS001-TCTGGGA	TCTGGGA	trt_sh.css	A375	96 h	1 µL	0.4874
174	DOS014_VCAP_6H:BRD-K22664171-001-01-3:4.95	BRD-K22664171	BRD-K22664171	trt_cp	VCAP	6 h	5 µM	0.4873
175	KDA005_A549_96H:TRCN0000154657:-666	TRCN0000154657	KEAP1	trt_sh	A549	96 h	1.5 µL	0.4872
176	CSS001_A375_96H:ATCATTG:1	CSS001-ATCATTG	ATCATTG	trt_sh.css	A375	96 h	1 µL	0.4871
177	DOS001_VCAP_24H:BRD-K97073065-001-01-6:5.12	BRD-K97073065	BRD-K97073065	trt_cp	VCAP	24 h	5 µM	0.4871
178	OEB006_VCAP_96H:BRDN0000461161:-666	ccsbBroad304_00851	IL6R	trt_oe	VCAP	96 h	3 µL	0.4869
179	CSS001_HT29_96H:CACAGAA:1	CSS001-CACAGAA	CACAGAA	trt_sh.css	HT29	96 h	1 µL	0.4868
180	AML001_U937_6H:BRD-K53325797:0.37037	BRD-K53325797	BRD-K53325797	trt_cp	U937	6 h	500 nM	0.4867
181	NMH001_NPC_6H:BRD-K56994829-001-01-9:10	BRD-K56994829	BRD-K56994829	trt_cp	NPC	6 h	10 µM	0.4865
182	KDB004_A375_96H:TRCN0000038998:-666	TRCN0000038998	CRYZ	trt_sh	A375	96 h	1 µL	0.486
183	CGS001_A549_96H:ZBTB20:1	CGS001-26137	ZBTB20	trt_sh.cgs	A549	96 h	1 µL	0.4859
184	DOS029_VCAP_24H:BRD-K05415698-001-01-5:5.03	BRD-K05415698	BRD-K05415698	trt_cp	VCAP	24 h	5 µM	0.4859
185	CGS001_HT29_96H:IGFBP2:1	CGS001-3485	IGFBP2	trt_sh.cgs	HT29	96 h	1 µL	0.4858
186	CGS001_HEPG2_96H:UBE2D1:1.5	CGS001-7321	UBE2D1	trt_sh.cgs	HEPG2	96 h	1.5 µL	0.4856
187	CSS001_HCC515_96H:ACCTCCG:2	CSS001-ACCTCCG	ACCTCCG	trt_sh.css	HCC515	96 h	2 µL	0.4856
188	KDA005_HCC515_96H:TRCN0000004266:-666	TRCN0000004266	STAT1	trt_sh	HCC515	96 h	2 µL	0.4852
189	CGS001_A549_96H:NOX1:1	CGS001-27035	NOX1	trt_sh.cgs	A549	96 h	1 µL	0.4851
190	KDB007_A549_96H:TRCN0000049361:-666	TRCN0000049361	GNPDA1	trt_sh	A549	96 h	1.5 µL	0.4851
191	CGS001_PC3_96H:RHOA:2	CGS001-387	RHOA	trt_sh.cgs	PC3	96 h	2 µL	0.485
192	CGS001_VCAP_120H:ZFP36L2:5	CGS001-678	ZFP36L2	trt_sh.cgs	VCAP	120 h	5 µL	0.4848
193	KDB005_PC3_96H:TRCN0000000465:-666	TRCN0000000465	BUB1B	trt_sh	PC3	96 h	2 µL	0.4847
194	CSS001_A375_96H:ACTTCAG:1	CSS001-ACTTCAG	ACTTCAG	trt_sh.css	A375	96 h	1 µL	0.4844
195	CGS001_HCC515_96H:CTBP2:2	CGS001-1488	CTBP2	trt_sh.cgs	HCC515	96 h	2 µL	0.4842
196	DOS056_MCF7_24H:BRD-K43433419-001-01-0:5	BRD-K43433419	BRD-K43433419	trt_cp	MCF7	24 h	5 µM	0.4841
197	KDA006_PC3_96H:TRCN0000018860:-666	TRCN0000018860	RPA3	trt_sh	PC3	96 h	2 µL	0.4841
198	CPC014_NPC_24H:BRD-U00779237-000-01-7:10	BRD-U00779237	WZ-3105	trt_cp	NPC	24 h	10 µM	0.4835
199	KDA009_HA1E_96H:TRCN0000219679:-666	TRCN0000219679	FGFR2	trt_sh	HA1E	96 h	2 µL	0.4834
200	CGS001_A549_96H:GART:1	CGS001-2618	GART	trt_sh.cgs	A549	96 h	1 µL	0.4833

Legend: the ranked highest connectivity score for mimicking the effect is presented with identification of the signature in the database (sig_id), identification of the chemical or genetic manipulation (pert_id), frequent name of the perturbagen (pert_iname), type of perturbation (pert_type), cell type (cell_id), duration (pert_itime), and dose of treatment (pert_idose) and the respective connectivity score (score). More information can be obtained on http://apps.lincscloud.org.